ADAPTIVE BAYESIAN DENSITY ESTIMATION USING PITMAN-YOR OR NORMALIZED INVERSE-GAUSSIAN PROCESS KERNEL MIXTURES

By Catia Scricciolo

Bocconi University

We consider Bayesian nonparametric density estimation using a Pitman-Yor or a normalized inverse-Gaussian process kernel mixture as the prior distribution for a density. The procedure is studied from a frequentist perspective. Using the stick-breaking representation of the Pitman-Yor process or the expression of the finite-dimensional distributions for the normalized-inverse Gaussian process, we prove that, when the data are replicates from an infinitely smooth density, the posterior distribution concentrates on any shrinking L^p-norm ball, $1 \leq p \leq \infty$, around the sampling density at a *nearly parametric* rate, up to a logarithmic factor. The resulting hierarchical Bayesian procedure, with a fixed prior, is thus shown to be adaptive to the infinite degree of smoothness of the sampling density.

*This research was mainly supported by grants from Luigi Bocconi University and the Italian Ministry of Education, University and Research (MIUR Grant No. 2008MK3AFZ_003), research project entitled "Bayesian Methods: Theoretical Developments and Novel Applications".

AMS 2000 subject classifications: Primary 62G07, 62G20

Keywords and phrases: Adaptation, Dirichlet process, nonparametric density estimation, nonparametric regression, normalized inverse-Gaussian process, Pitman-Yor process, posterior distribution, rate of convergence.

1. Introduction. Consider the problem of estimating a density on the real line from independent and identically distributed (i.i.d.) observations, taking a Bayesian nonparametric approach. A prior is defined on a metric space of probability measures with Lebesgue density and a summary of the posterior, typically the posterior expected density, can be employed as an estimator. Since the seminal articles of Ferguson [6] and Lo [22], the idea of constructing priors on spaces of densities by convolving a fixed kernel with a random distribution has been successfully exploited in density estimation: a kernel mixture may provide an efficient approximation scheme, possibly resulting in a minimax-optimal (up to a logarithmic factor) speed of concentration for the posterior on shrinking balls around the sampling density.

Recent literature on Bayesian kernel density estimation has mainly focussed on posterior contraction rates relative to the Hellinger or the L^1-metric, using a Dirichlet process mixture of normals (or generalized normals). Ghosal and van der Vaart [10] found a nearly parametric rate for estimating *infinitely smooth* densities that are in the model, *i.e.*, are themselves normal mixtures, while Shen and Ghosal [31], extending the result of Kruijer *et al.* [20] on finite Dirichlet mixtures, have proved that fully rate-adaptive multivariate density estimation of *ordinary smooth* densities over Hölder regularity scales can be performed using *infinite* Dirichlet mixtures of Gaussians, without bandwidth shrinkage in the prior for the scale.

Even if much progress has been done during the last decade in understanding frequentist asymptotic properties of kernel mixture models for Bayesian density estimation, there seems to be a lack of results concerning adaptive estimation of infinitely smooth densities with respect to more general loss-functions than the Hellinger metric, using other processes, apart from the Dirichlet process, as priors for the mixing. In this article, we investigate the question of how to complement and generalize existing results on posterior contraction rates by considering adaptive estimation of infinitely smooth densities using, *e.g.*, either the Pitman-Yor or the normalized inverse-Gaussian process as priors for the mixing distribution of general kernel mixtures.

We prove that for densities with Fourier transform satisfying an exponential moment condition, virtually implying that the characteristic function decreases at worst at an exponential power rate, an almost parametric rate of posterior contraction arises in all L^p-norms, $1 \leq p \leq \infty$, under different

priors (possibly) affecting only the power of the logarithm term which, admittedly, may not be optimal. In fact, it is known that the minimax rate for estimating an entire density f such that $\sup_x |f(x+iy)| \leq M \exp\{c|y|^\rho\}$, with $\rho > 1$, is $n^{-1/2}(\log n)^{(\rho-1)/(2\rho)}$ in all L^p-norm, $2 \leq p < \infty$, see, *e.g.*, Theorem 4.1 in Ibragimov [16], page 366, and the reference therein. Such a fast rate is roughly explainable from the fact that the spaces of analytic functions are only slightly bigger than the finite-dimensional spaces in terms of metric entropy.

Such results are of interest for a variety of reasons: they may constitute a first step, beyond the Dirichlet process, towards the study of posterior contraction rates for more involved process priors recently proposed in the literature. Also, they provide an indication on the performance of Bayes' procedures for adaptive estimation of densities belonging to a class, extensively considered in the frequentist literature for nonparametric curve estimation, which provides an alternative to classes of densities that are only finitely smooth.

The main challenge in extending the adaptation result from the ordinary to the infinitely smooth case rests in finding a finite mixing distribution, with a sufficiently restricted number of support points, such that the corresponding Gaussian mixture approximates the sampling density, in Kullback-Leibler divergence, with an exponentially small error in terms of the inverse of the bandwidth. Such a finitely supported mixing distribution may be found by matching the moments of an *ad hoc* constructed mixing density, for which, however, the twicing kernel method used by Kruijer *et al.* [20] does not seem to be well-suited because of the infinite degree of smoothness of the true density. There seems to be limitations implicitly coming from the kernel which are by-passed using superkernels, *e.g.*, the sinc kernel, whose usefulness in density estimation has been pointed out by, among others, Devroye [5]. The crux and a main contribution of this article is the development of an approximation result for analytic densities with exponentially decaying Fourier transforms, cf. Lemma 8.1. We believe this result can be of autonomous interest as well and possibly exploited by frequentist methods in adaptive density estimation for clustering with Gaussian mixtures, along the lines of Maugis and Michel [23].

When assessing posterior rates, a major difficulty is the evaluation of the prior concentration rate, calculated bounding below the prior probability of Kullback-Leibler type neighbourhoods by the prior probability of an L^1-ball of the right dimension. For the normalized inverse-Gaussian process, the expression of the finite-dimensional distributions is used to estimate the probability of an L^1-ball as done in the literature for the Dirichlet process.

For the Pitman-Yor process, instead, we exploit the stick-breaking representation to obtain lower bounds on the probabilities of L^1-balls of the mixing weights and the locations. We expect that this technique can be applied to other stick-breaking processes in future papers.

For the sake of clarity, the exposition is focussed on density estimation, but other statistical settings are implicitly covered: for example, fixed design linear regression with unknown error distribution, as described in Ghosal and van der Vaart [11], pages 205–206. Extension of these results to a multivariate setting seems imminent along the lines of Shen and Ghosal [31] and is not pursued here.

The organization of the article is as follows. In Section 2, we fix the notation and review preliminary definitions. In Section 3, we state a result on posterior rates for general kernel mixtures, highlighting the connection with rates of contraction of the posterior for the mixing distribution (proofs are postponed to Section 5). Main results are reported in Section 4, where after investigating the achievability of the error rate $1/\sqrt{n}$, up to a logarithmic factor, for *super-smooth* densities that are in the model, which helps developing mathematical tools, we focus on adaptive estimation of *analytic* densities using infinite Gaussian mixtures. Prior estimates are given in Section 6. Section 7 and Section 8 report the proofs of the main theorems. Auxiliary results are collected in the Appendix.

2. Notation and preliminaries. In this section, we introduce some notation and review definitions used throughout the article. The model is assumed to be a *location* mixture,

$$f_{F,\sigma}(x) := (F * K_\sigma)(x) = \int_{\mathbb{R}} \sigma^{-1} K((x-\theta)/\sigma)\,\mathrm{d}F(\theta), \qquad x \in \mathbb{R},$$

where K denotes the kernel density, σ the scale parameter and F the mixing distribution. Kernels herein considered are characterized via a condition involving the Fourier transform. Such a condition is hereafter stated for a generic probability density function. Let $\hat{f}(t) := \int_{\mathbb{R}} e^{itx} f(x)\,\mathrm{d}x$, $t \in \mathbb{R}$, be the Fourier transform or characteristic function of f. We say that f is *super-smooth* if its characteristic function satisfies the following condition: for constants ρ, $r > 0$ and $0 < L < \infty$,

$$I_{\rho,r}(f) := \int_{\mathbb{R}} |\hat{f}(t)|^2 \exp\{2(\rho|t|)^r\}\,\mathrm{d}t \le 2\pi L. \tag{2.1}$$

Condition (2.1) implies that the behaviour of $|\hat{f}|$ is virtually described by $\exp\{-(\rho|t|)^r\}$ as $|t| \to \infty$. Densities with Fourier transforms satisfying requirement (2.1) are *infinitely* differentiable on $\mathbb{R}$, see, *e.g.*, Theorem 11.6.2.

in Kawata [19], pages 438–439. Also, they are bounded. Set $C(\rho, r) := \int_0^\infty \exp\{-2(\rho t)^r\}\,\mathrm{d}t = (2\rho^r)^{-1/r}\Gamma(1+1/r)$,

$$\|f\|_\infty \le (2\pi)^{-1}\int_{\mathbb{R}} |\hat{f}(t)|\,\mathrm{d}t \le L + \pi^{-1}C(\rho, r) < \infty, \tag{2.2}$$

cf. Lemma 1 in Butucea and Tsybakov [3], page 35. Super-smooth densities constitute a somewhat larger class than that of analytic densities, including important examples like Gaussian, Cauchy, general symmetric stable laws, Student's-t, distributions with characteristic functions vanishing outside a compact, as well as their mixtures and convolutions.

EXAMPLE 2.1. *Symmetric stable laws*, which have characteristic functions of the form $e^{-(\rho|t|)^r}$, $t \in \mathbb{R}$, for some $\rho \ge 0$ and $0 < r \le 2$, are super-smooth. Here, r is called the *index* of the stable law, except in the degenerate case $\rho = 0$ where the Fourier transform is identically equal to 1 and the law is a point mass at 0. We rule out this case. Cauchy laws Cauchy$(0, \sigma)$ are stable with $r = 1$ and $\rho = \sigma$. Normal laws $\mathrm{N}(0, \sigma^2)$ are stable with $r = 2$ and $\rho = \sigma/\sqrt{2}$.

EXAMPLE 2.2. *Student's-t distribution* with $\nu > 0$ degrees of freedom has characteristic function verifying (2.1) for $r = 1$:

$$\widehat{f}_{t_\nu}(t) \cong \frac{\sqrt{\pi}}{\Gamma(\nu/2)2^{(\nu-1)/2}}(\sqrt{\nu}|t|)^{(\nu-1)/2}e^{-\sqrt{\nu}|t|} \qquad \text{as } |t| \to \infty, \tag{2.3}$$

see formula (4.8) in Hurst [15], page 5.

EXAMPLE 2.3. *Exponential Power Distributions* (EPD's) with shape parameter p that is an *even* integer have characteristic functions satisfying (2.1). A random variable X has an EPD with *location* parameter $\theta = \mathrm{E}[X]$, *shape* parameter (or *exponent*) p and *scale* parameter $\sigma \equiv \sigma_p = \{\mathrm{E}[|X-\theta|^p]\}^{1/p}$, in symbols, $X \sim \mathrm{EPD}(\theta, \sigma, p)$, with $\theta \in \mathbb{R}$ and $\sigma, p > 0$, if it has density $f_{\theta,\sigma,p}(x) = [2\sigma p^{1/p}\Gamma(1+1/p)]^{-1}\exp\{-(|x-\theta|/\sigma)^p/p\}$, $x \in \mathbb{R}$. It is known from Pogány and Nadarajah [29], page 205, that, for $p > 1$,

$$\widehat{f_{\theta,\sigma,p}}(t) = \frac{\sqrt{\pi}e^{it\theta}}{\Gamma(1/p)}\sum_{k=0}^\infty \frac{\Gamma((2k+1)/p)}{\Gamma((2k+1)/2)} \times \frac{[-(\sigma t p^{1/p}/2)^2]^k}{k!}, \qquad t \in \mathbb{R}, \tag{2.4}$$

where the series converges for $p > 1$. Proposition A.1 asserts that, when $p = 2m$, $m \in \mathbb{N}$, $\widehat{f_{\theta,\sigma,2m}}(t) \lesssim e^{-ct^2}$, $t \in \mathbb{R}$, thus, the corresponding distribution function is analytic.

EXAMPLE 2.4. *Densities with characteristic functions vanishing outside a symmetric convex compact set* are super-smooth. Let Λ be a symmetric convex compact set in $\mathbb{R}^k$, $k \geq 1$. Denote by Σ_Λ the class of densities with characteristic functions equal to 0 outside Λ. The set Σ_Λ is essentially infinite-dimensional, nevertheless, for $p \geq 2$, $\inf_{f_n} \sup_{f \in \Sigma_\Lambda} \mathrm{E}_f^n[\|f_n - f\|_p^s] \leq c_s n^{-s/2}$. Moreover, for $p = s = 2$, the precise asymptotic bound holds:

$$\lim_{n\to\infty} n \inf_{f_n} \sup_{f\in\Sigma_\Lambda} \mathrm{E}_f^n[\|f_n - f\|_2^2] = \frac{\mathrm{meas}(\Lambda)}{(2\pi)^k},$$

see Hasminskii and Ibragimov [14], page 1008, and the references therein. Let $\Lambda = [-T, T]$, for $0 < T < \infty$. For any $f \in \Sigma_\Lambda$, it is $I_{\rho, r}(f) \leq 2\pi L$ for every ρ, $r > 0$ and $L \geq \pi^{-1} T \exp\{2(\rho T)^r\}$. The *Fejér-de la Vallée-Poussin* density $f(x) = (2\pi)^{-1}[(x/2)^{-1}\sin(x/2)]^2$, $x \in \mathbb{R}$, having $\hat{f}(t) = (1 - |t|)^+$, $t \in \mathbb{R}$, is the typical example of density in Σ_Λ, for $\Lambda = [-1, 1]$.

Given the model $f_{F,\sigma} = F * K_\sigma$, a prior is constructed on the space of Lebesgue univariate densities by putting priors on the scale and the mixing distribution. The scale parameter is assumed to be distributed, independently of F, according to a distribution G on $(0, \infty)$. Let $\Pi \otimes G$ denote the overall prior on $\mathscr{M}(\mathbb{R}) \times (0, \infty)$, where $\mathscr{M}(\mathbb{R})$ stands for the set of probability measures on $\mathbb{R}$. Then, $\Pi \otimes G$ induces a prior on $\mathscr{F} := \{f_{F,\sigma} : (F, \sigma) \in \mathscr{M}(\mathbb{R}) \times (0, \infty)\}$ via the mapping $(F, \sigma) \mapsto f_{F,\sigma}$. We assume that $\mathscr{F}$ is equipped with a metric d, either the Hellinger $d_{\mathrm{H}}(f, g) := (\int (f^{1/2} - g^{1/2})^2 \,\mathrm{d}\lambda)^{1/2}$, where λ denotes Lebesgue measure on $\mathbb{R}$, or the one induced by the L^p-norm, $\|f - g\|_p := (\int |f - g|^p \,\mathrm{d}\lambda)^{1/p}$, $1 \leq p < \infty$, the sup-norm being defined as $\|f - g\|_\infty := \sup_{x\in\mathbb{R}} |f(x) - g(x)|$. We study rates of contraction for the posterior of $\Pi \otimes G$, assuming that $X^{(n)} := (X_1, \ldots, X_n)$ are i.i.d. observations from f_0, which may or may not be itself a kernel mixture. A sequence $\delta_n \to 0$, as $n \to \infty$, is said to be an *upper bound* on the posterior rate of convergence relative to a metric d if, for some constant $0 < M < \infty$, $(\Pi \otimes G)(\{f_{F,\sigma} : d(f_{F,\sigma}, f_0) \geq M\delta_n\}|X^{(n)}) \to 0$, P_0^∞-almost surely or in P_0^n-probability, where P_0 stands for the true probability measure. In the next section, we provide a result on posterior rates of contraction for mixture models with super-smooth kernels.

3. Posterior rates of contraction for kernel mixtures. We derive a theorem for assessing rates of posterior contraction for kernel mixtures in terms of the prior concentration rate only. The assertion yields minimax-optimal rates, up to a log-factor, for every L^p-norm, $2 \leq p \leq \infty$, when the prior concentration rate is nearly parametric. Moreover, it relaxes the

condition on the prior for the scale to a condition involving only the lower tail by requiring an exponential decay at zero.

(P) The prior distribution G for σ satisfies the tail condition $G(\sigma) \lesssim e^{-d\sigma^{-\gamma}}$ as $\sigma \to 0$, for some constants $d > 0$ and $1 < \gamma \leq \infty$.

The assertion of the theorem holds for any density that is well approximated by its convolution with the sinc kernel

$$\operatorname{sinc}(x) := \begin{cases} \dfrac{\sin x}{\pi x}, & \text{if } \ x \neq 0, \\ \dfrac{1}{\pi}, & \text{if } \ x = 0. \end{cases}$$

This is an unconventional kernel, in the sense that it may take negative values. It is Riemann integrable, $\int \operatorname{sinc} \, \mathrm{d}\lambda = 1$, but not Lebesgue integrable, $\operatorname{sinc} \notin L^1(\mathbb{R})$. Also, $\widehat{\operatorname{sinc}}(t) = 1_{[-1,\,1]}(t)$, $t \in \mathbb{R}$. To state the theorem, let $B_{\mathrm{KL}}(f_0;\, \varepsilon_n^2) := \{f_{F,\,\sigma} :\ \mathrm{E}[\log(f_{F,\,\sigma}/f_0)] \leq \varepsilon_n^2,\, \mathrm{E}[(\log(f_{F,\,\sigma}/f_0))^2] \leq \varepsilon_n^2\}$.

THEOREM 3.1. *Let K be a probability density with characteristic function satisfying* (2.1) *for some constants $\rho, r > 0$ and $0 < L < \infty$. Let ε_n be a sequence such that $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$ as $n \to \infty$. For every $2 \leq p \leq \infty$, define $\delta_n := \varepsilon_n(n\varepsilon_n^2)^{(1-1/p)/2}$. Suppose that, for $f_0 \in L^p(\mathbb{R})$ such that $\|f_0 * \operatorname{sinc}_{2^{-J_n}} - f_0\|_p = O(\delta_n)$, with $2^{J_n} = O(n\varepsilon_n^2)$,*

$$(\Pi \otimes G)(B_{\mathrm{KL}}(f_0;\, \varepsilon_n^2)) \gtrsim \exp\{-Cn\varepsilon_n^2\} \qquad \text{for some constant } C > 0, \tag{3.1}$$

where Π is a prior for F and G a prior for σ satisfying assumption (P), *with $\gamma > 1$ such that $n\varepsilon_n^2 \gtrsim (\log n)^{1/[r(1-1/\gamma)]}$. If $\delta_n \to 0$ as $n \to \infty$, then there exists a constant $0 < M < \infty$ such that*

$$(\Pi \otimes G)(\{f_{F,\,\sigma} :\ \|f_{F,\,\sigma} - f_0\|_p \geq M\delta_n\} | X^{(n)}) \to 0 \qquad \text{in } P_0^n\text{-probability.}$$

When the employed kernel has characteristic function decreasing at an exponential power rate and f_0 is a kernel mixture with compactly supported mixing distribution, the preceding theorem yields rates of contraction, relative to the Wasserstein distance of order 1, for the posterior of the mixing distribution. Let $(\Theta,\, d)$, $\Theta \subseteq \mathbb{R}$, be a measurable metric space with the Borel σ-field. For $p \geq 1$, define the *Wasserstein distance of order p* between any two Borel probability measures μ and ν on Θ with finite pth-moment (*i.e.*, $\int_\Theta d^p(x,\, x_0)\, \mathrm{d}\mu(x) < \infty$ for some (and hence any) x_0 in Θ) as $W_p(\mu,\, \nu) := (\inf_{\gamma \in \Gamma(\mu,\, \nu)} \int_{\Theta \times \Theta} d^p(x,\, y)\, \mathrm{d}\gamma(x,\, y))^{1/p}$, where γ runs over the set of all joint probability measures on $\Theta \times \Theta$ with marginal distributions

μ and ν. When $p = 2$, we take d to be the Euclidean distance on Θ. Let $\mathrm{diam}(\Theta) := \sup\{d(x,\, y) : x,\, y \in \Theta\}$ be the diameter of Θ. From the definition, $W_p(\mu,\, \nu) \in [0,\, \mathrm{diam}(\Theta)]$. If Θ is compact, then $\mathrm{diam}(\Theta) < \infty$.

COROLLARY 3.1. *Let K be a symmetric density around 0, with characteristic function satisfying (2.1) for some constants $\rho,\, r > 0$ and $0 < L < \infty$ and, furthermore,*

$$\text{for some constants } B,\, \beta > 0, \qquad |\hat{K}(t)| \geq B \exp\{-(\beta|t|)^r\}, \qquad t \in \mathbb{R}. \tag{3.2}$$

Let $f_0 = f_{F_0,\,1}$, with F_0 supported on a compact set $\Theta \subset \mathbb{R}$. Let Π be a prior on $\mathscr{M}(\Theta)$. If condition (3.1) is verified for a sequence ε_n such that $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$ as $n \to \infty$, then, for a sufficiently large constant $0 < M < \infty$,

$$\Pi(\{F : W_1(F,\, F_0) < M(\log n)^{-1/r}\}|X^{(n)}) \to 1 \qquad \textit{in } P_0^n\textit{-probability.}$$

4. Posterior rates of contraction for specific priors on the mixing.

In this section, we derive rates of contraction for specific priors on the mixing distribution, *i.e.*, the Pitman-Yor process, which renders the Dirichlet process as a special case, and the normalized inverse-Gaussian process. Nearly parametric rates arise either when the sampling density is within the model or when it is a generic analytic density.

4.1. *Estimation of densities with kernel mixture representation.*

We begin the analysis from the case where f_0 is itself a kernel mixture,

$$f_0(x) = f_{F_0,\,\sigma_0}(x) = (F_0 * K_{\sigma_0})(x), \qquad x \in \mathbb{R},$$

where F_0 and σ_0 denote the true values of the mixing distribution and the scale parameter, respectively. Results are obtained under the following assumptions.

(A_0) The kernel density $K : \mathbb{R} \to \mathbb{R}^+$ is symmetric around 0, monotone decreasing in $|x|$ and satisfies the tail condition $K(x) \gtrsim e^{-c|x|^\kappa}$ for large $|x|$, for some constants $c > 0$ and $0 < \kappa < \infty$.

(A_1) The true mixing distribution F_0 satisfies the tail condition

$$F_0(\{\theta : |\theta| > t\}) \lesssim e^{-c_0 t^{\varpi}} \qquad \text{for large } t > 0, \tag{4.1}$$

for some constants $c_0 > 0$ and $0 < \varpi \leq \infty$.

(A_2) The base measure α has a continuous and positive Lebesgue density α' such that, for some constants $b > 0$ and $0 < \delta \leq \infty$, satisfies the tail condition

$$\alpha'(\theta) \propto e^{-b|\theta|^\delta} \qquad \text{for large } |\theta|. \tag{4.2}$$

(A_3) The prior distribution G for σ has a continuous and positive Lebesgue density on an interval containing σ_0 and, for constants $d > 0$, $1 < \gamma \leq \infty$ and $0 < \varrho \leq \infty$, satisfies $G(\sigma) \lesssim e^{-d\sigma^{-\gamma}}$ as $\sigma \to 0$ and $1 - G(\sigma) \lesssim \sigma^{-\varrho}$ as $\sigma \to \infty$.

The case where $\varpi = \infty$ in (A_1) corresponds to a compactly supported mixing distribution. The same holds for the base measure α when $\delta = \infty$ in (A_2). As far as assumption (A_3) is concerned, an inverse-gamma distribution on σ^2 is an eligible prior, in fact, for a suitable $d > 0$, we have $G(\sigma) \lesssim e^{-d\sigma^{-2}}$.

Stick-breaking processes and the Pitman-Yor process. We consider the class of stick-breaking processes which includes, as important special cases, the Dirichlet process, the two-parameter Poisson-Dirichlet process or Pitman-Yor process, see Pitman and Yor [27], the beta two-parameter process, see Ishwaran and Zarepour [17], Ishwaran and James [18], and the geometric stick-breaking process, see Mena *et al.* [24]. The trajectories of a stick-breaking process F can be (almost surely) represented as $F = \sum_{j=1}^{\infty} W_j \delta_{Z_j}$, where δ_{Z_j} denotes a point mass at Z_j. The random variables Z_j, $j \in \mathbb{N}$, are i.i.d. $\bar{\alpha}$, where $\bar{\alpha}$ is a non-atomic (*i.e.*, $\bar{\alpha}(\{z\}) = 0$ for every $z \in \mathbb{R}$) probability measure over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined as $\bar{\alpha} := \alpha/\alpha(\mathbb{R})$, α being a positive and finite measure. The random variables W_j, $j \in \mathbb{N}$, are independent of the Z_j's and such that $0 \leq W_j \leq 1$, with $\sum_{j=1}^{\infty} W_j \overset{\text{a.s.}}{=} 1$. Also,

$$W_1 = V_1, \qquad W_j = V_j \prod_{h=1}^{j-1} (1 - V_h), \qquad j \geq 2, \tag{4.3}$$

with $V_j | H_j \overset{\text{indep}}{\sim} H_j$, where H_j is a probability measure on $[0,\,1]$. A necessary and sufficient condition for $\sum_{j=1}^{\infty} W_j \overset{\text{a.s.}}{=} 1$ to hold is that $\sum_{j=1}^{\infty} \log(1 - \mathrm{E}_{H_j}[V_j]) = -\infty$, see, *e.g.*, Lemma 1 in Ishwaran and James [18], pages 162 and 170. Consider a stick-breaking process where, for $0 \leq d < 1$ and $c > -d$, $V_j \overset{\text{indep}}{\sim} \mathrm{Beta}(1-d,\, c+dj)$, $j \in \mathbb{N}$. The parameter c is called the *concentration parameter* and d the *discount parameter*. The resulting process is called the *two-parameter Poisson-Dirichlet process* or *Pitman-Yor process* with parameters c, d and *base measure* $\bar{\alpha}$, denoted $F \sim \mathrm{PY}(c,\, d,\, \bar{\alpha})$:

$$\begin{aligned} F \quad &\sim \quad \sum_{j=1}^{\infty} \left[V_j \prod_{h=1}^{j-1} (1 - V_h) \right] \delta_{Z_j} \\ V_j \quad &\overset{\text{indep}}{\sim} \quad \mathrm{Beta}(1 - d,\, c + dj) \\ Z_j \quad &\overset{\text{iid}}{\sim} \quad \bar{\alpha}. \end{aligned}$$

The case where $d = 0$ and $c = \alpha(\mathbb{R})$ returns the Dirichlet process $\mathrm{D}(\alpha)$. Note that, in both the Dirichlet process and the Pitman-Yor process, the weights $\{V_j \prod_{h=1}^{j-1}(1 - V_h)\}_{j\geq 1}$ are the weights of the process in size-biased order. When $c = 0$, the Pitman-Yor process reduces to a stable process. When $c = 0$ and $d = 1/2$, the stable process is a normalized inverse-gamma process. There are no known analytic expressions for its finite-dimensional distributions, except for the cases where $d = 0$ and $d = 1/2$. Indeed, the Dirichlet process, the Pitman-Yor process, with $d = 1/2$, and the normalized inverse-Gaussian process are the only known priors for which explicit expressions of the finite-dimensional distributions are available.

In order to state the main result of the section, for $\kappa,\, r > 0$, let ϖ be such that

$$\max\{\kappa,\, [1 + I_{(1,\,\infty)}(r)/(r - 1)]\} \leq \varpi \leq \infty \tag{4.4}$$

and let τ be defined as

$$\tau := 1 + \big[1/r - \big(1 - I_{(0,\,\infty)}(\varpi)/\varpi\big)\big]\, I_{(0,\,1]}(r)/2. \tag{4.5}$$

THEOREM 4.1. *Let K be as in assumption* (A_0)*, with characteristic function satisfying* (2.1) *for some constants $\rho,\, r > 0$ and $0 < L < \infty$. Let $f_0 = F_0 * K_{\sigma_0}$, with*

(*i*) *F_0 satisfying* (A_1) *for some constants $c_0 > 0$ and ϖ as in* (4.4)*.*

Let the prior for F be a $\mathrm{PY}(c,\, d,\, \bar{\alpha})$, with $c > -d$ and $0 \leq d < 1$. Assume that

(*ii*) *α satisfies* (A_2) *for some constants $b > 0$ and $0 < \delta < \infty$, with $\delta \leq \varpi$ whenever $\varpi < \infty$,*

(*iii*) *G satisfies* (A_3) *for $\gamma > 1$ and $\gamma \geq \{1 - \{2r[\tau + (\tau - 1/2)I_{(0,\,\infty)}(d)]\}^{-1}\}^{-1}$, with τ as in* (4.5)*,*

then the posterior rate of convergence relative to the L^p-norm, $1 \leq p \leq \infty$, is $n^{-1/2}(\log n)^{\mu}$ for a suitable $\mu > 0$.

Normalized inverse-Gaussian process. We start by recalling the definition of the normalized inverse-Gaussian (N-IG) distribution. The random vector $(Z_1,\, \ldots,\, Z_N)$, $N \geq 2$, is said to have a N-IG distribution with parameters $(\alpha_1,\, \ldots,\, \alpha_N)$, $\alpha_j \geq 0$ for every $j = 1,\, \ldots,\, N$ and $\alpha_j > 0$ for at least one j, denoted $\text{N-IG}(\alpha_1,\, \ldots,\, \alpha_N)$, if it has probability density function over the unit $(N-1)$-simplex Δ^{N-1}

$$f(z_1,\, \ldots,\, z_{N-1}) \;=\; \frac{e^{\sum_{j=1}^N \alpha_j} \prod_{j=1}^N \alpha_j}{2^{N/2-1}\pi^{N/2}} \times K_{-N/2}(\sqrt{\mathcal{A}_N(z_1,\, \ldots,\, z_{N-1})})$$

$$\times\,(\mathcal{A}_N(z_1,\,\ldots,\,z_{N-1}))^{-N/4}$$
$$\times\,[z_1\times\ldots\times z_{N-1}\times(1-z_1-\,\ldots\,-z_{N-1})]^{-3/2}$$
$$=:\ \prod_{r=1}^{4} h_r(z_1,\,\ldots,\,z_{N-1}), \tag{4.6}$$

where $K_{-N/2}(\cdot)$ is the modified Bessel function of the second kind (see, *e.g.*, Abramowitz and Stegun [1], Ch. 9) and $\mathcal{A}_N(z_1,\,\ldots,\,z_{N-1}):=\sum_{j=1}^{N-1}(\alpha_j^2/z_j)+\alpha_N^2/(1-\sum_{j=1}^{N-1}z_j)$.

Consider a space $\mathbb{X}$ with a σ-algebra $\mathcal{A}$ of subsets of $\mathbb{X}$. Let α be a finite and positive measure on $(\mathbb{X},\,\mathcal{A})$. Following Lijoi *et al.* [21], a random probability measure F is called a *normalized inverse-Gaussian process* on $(\mathbb{X},\,\mathcal{A})$, with parameter α, denoted N-IG(α), if, for every finite measurable partition $A_1,\,\ldots,\,A_N$ of $\mathbb{X}$, the probability vector $(F(A_1),\,\ldots,\,F(A_N))$ has a N-IG distribution with parameters $(\alpha(A_1),\,\ldots,\,\alpha(A_N))$. The N-IG process has wider support around its modes than the Dirichlet process.

THEOREM 4.2. *Assume the set-up and conditions of Theorem 4.1 except that the prior for F is a* N-IG(α)*, where α is a finite and positive measure on $\mathbb{R}$. Then, the posterior rate of convergence relative to the L^p-norm, $1\le p\le\infty$, is $n^{-1/2}(\log n)^{\mu}$ for a suitable $\mu>0$.*

The proof is the same as for the Dirichlet process and is omitted.

4.2. *Adaptive estimation of analytic densities.* In this section, we study adaptive estimation of analytic densities using infinite Gaussian mixtures. We assume that f_0 satisfies the following conditions, where we denote by $C^{\omega}(\mathbb{R})$ the class of analytic functions on $\mathbb{R}$.

(a) *Smoothness*: $f_0\in C^{\omega}(\mathbb{R})$ is a probability density with characteristic function satisfying (2.1) for some constants $\rho_0>0$, $0<r_0<2$ and $0<L_0<\infty$. Furthermore, $\log f_0$ is locally Lipschitz continuous:

$$\exists\,\delta>0 \text{ s.t. } \forall\,x,\,y:\ |y-x|\le\delta,\qquad |\log f_0(y)-\log f_0(x)|\le Q(x)|y-x|,$$

where Q is a non-negative polynomial function of degree at most $(q-1)$, $q\ge 2$, such that $\mathrm{E}_0[\exp\{[Q(X)]^{q/(q-1)}\}]<\infty$;

(b) *Monotonicity*: f_0 is a strictly positive and bounded density, non-decreasing on $(-\infty,\,a)$, non-increasing on $(b,\,\infty)$ and such that $f_0(x)\ge\ell_0>0$ on $[a,\,b]$;

(c) *Tails*: there exists a constant $M_0>0$ such that $f_0(x)\le M_0\phi(x)$ for all $x\in\mathbb{R}$.

Conditions (a)–(c) are satisfied, for example, by EPD's with p that is an even integer. Even if the greatest value of r_0 verifying condition (2.1) may be greater or equal than 2, we can always take r_0 to be in $(0, 2)$. This will only cause a loss in the logarithmic term of the rate, not affecting the power of n.

THEOREM 4.3. *Suppose f_0 is a probability density satisfying (a)-(c) for $q \geq 2$, $0 < r_0 < q/(3q-2)$ and $r_0(q-1)/q < \beta < 1 - r_0/2$. Let the model be $f_{F,\sigma} = F * \phi_\sigma$, with $F \sim \mathrm{PY}(c,\, d,\, \bar{\alpha})$ for $c > -d$ and $0 \leq d < 1$. Alternatively, let $F \sim \text{N-IG}(\alpha)$. Assume that*

- *(i) α satisfies (A_2) for constants $b > 0$ and $0 < \delta \leq 2$,*
- *(ii) G satisfies (A_3) for some $\gamma = 2$ and $0 < \varrho \leq \infty$.*

The posterior rate of convergence relative to the L^p-norm, $1 \leq p \leq \infty$, is $\varepsilon_n = n^{-1/2}(\log n)^\mu$ for a suitable $\mu > 0$.

5. Proofs of Theorem 3.1 and of Corollary 3.1. The following lemma provides an upper bound on the L^p-approximation error of a density, whose Fourier transform either vanishes outside a compact or decays exponentially fast, by its convolution with the sinc kernel. In order to state it, we define, for any probability density f, the positive (possibly infinite) constant $S_f := \sup\{|t| : |\hat{f}(t)| \neq 0\}$. If

- $S_f < \infty$, then $\mathrm{supp}(\hat{f}) \subseteq (-S_f,\, S_f)$,
- $S_f = \infty$, then $\hat{f}(t) > 0$ for every $t \in \mathbb{R}$.

Next, we recall some basic facts on the Fourier transform. If $\int_{\mathbb{R}} |\hat{f}(t)|\,\mathrm{d}t < \infty$, then f can be recovered from $\hat{f}$ using the inversion formula $f(x) = (2\pi)^{-1}\int_{\mathbb{R}} e^{-itx}\hat{f}(t)\,\mathrm{d}t$, $x \in \mathbb{R}$. Furthermore, f is continuous and bounded, $\|f\|_\infty \leq (2\pi)^{-1}\int_{\mathbb{R}} |\hat{f}(t)|\,\mathrm{d}t < \infty$.

LEMMA 5.1. *Let f be a probability density on $\mathbb{R}$ with characteristic function satisfying (2.1) for some constants $\rho,\, r > 0$ and $0 < L < \infty$. For any fixed $\sigma > 0$,*

- *if* $S_f < 1/\sigma$, *then* $\|f * \mathrm{sinc}_\sigma - f\|_p = 0$, $\quad 1 \leq p \leq \infty$,
- *if* $S_f = \infty$, *then* $\|f * \mathrm{sinc}_\sigma - f\|_p \lesssim e^{-(\rho/\sigma)^r/2}$, $\quad 2 \leq p \leq \infty$.

PROOF. By the inversion formula for characteristic functions,

$$(f * \mathrm{sinc}_\sigma)(x) - f(x) = \frac{1}{2\pi}\int_{|t|>1/\sigma} e^{-itx}\hat{f}(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$$

If $S_f < 1/\sigma$, then $\int_{|t|>1/\sigma} e^{-itx}\hat{f}(t)\,\mathrm{d}t = 0$ identically and $\|f*\mathrm{sinc}_\sigma - f\|_p = 0$ for every $1 \le p \le \infty$. Next, suppose $S_f = \infty$. For any function $g \in L^p$, with $2 \le p < \infty$, we have $\|g\|_p^p \le C_p\|\hat{g}\|_q^q$, where $q^{-1} := (1-p^{-1}) \in [1/2, 1)$ and C_p is a constant depending only on p, see, *e.g.*, Theorem 74 in Titchmarsh [32], page 96. Hence, $\|f*\mathrm{sinc}_\sigma - f\|_p^p \le C_p\|\hat{f}(\widehat{\mathrm{sinc}_\sigma} - 1)\|_q^q = C_p\int_{|t|>1/\sigma}|\hat{f}(t)|^q\,\mathrm{d}t$. By the Cauchy-Schwarz inequality and the assumption that f satisfies (2.1),

$$\begin{aligned}
\int_{|t|>1/\sigma}|\hat{f}(t)|^q\,\mathrm{d}t &\le \int_{|t|>1/\sigma}|\hat{f}(t)|\,\mathrm{d}t \\
&\lesssim \left(\int_{|t|>1/\sigma}\exp\{-2(\rho|t|)^r\}\,\mathrm{d}t\right)^{1/2} \\
&\lesssim \sigma^{-(1-r)/2}e^{-(\rho/\sigma)^r} \lesssim e^{-(\rho/\sigma)^r/2},
\end{aligned} \tag{5.1}$$

where $\int_{1/\sigma}^{\infty}\exp\{-2(\rho t)^r\}\,\mathrm{d}t = r^{-1}(2\rho^r)^{-1/r}\Gamma(r^{-1}, 2(\rho/\sigma)^r)$, for $\Gamma(a, z) = \int_z^\infty t^{a-1}e^{-t}\,\mathrm{d}t$, with $a, z > 0$, the upper incomplete gamma function. It is known that $\Gamma(a, z) \sim z^{a-1}e^{-z}$ as $z \to \infty$. The case where $p = \infty$ is treated implicitly in (5.1). $\square$

The result can be extended to all L^p-metrics, $1 \le p \le \infty$, replacing the sinc kernel with a superkernel, which is absolutely integrable. By definition, a *superkernel* L is a symmetric, absolutely integrable function $\int|L|\,\mathrm{d}\lambda < \infty$, with $\int L\,\mathrm{d}\lambda = 1$, having an absolutely integrable Fourier transform $\hat{L}$ (hence L is bounded) with the properties that $\hat{L} = 1$ identically on $[-1, 1]$ and $|\hat{L}| < 1$ off $[-1, 1]$. The interval $[-1, 1]$ is chosen for convenience only, any neighborhood of the origin is fine. Superkernels necessarily have infinite support.

LEMMA 5.2. *Let f be a probability density on $\mathbb{R}$ with characteristic function satisfying (2.1) for some constants $\rho, r > 0$ and $0 < L_f < \infty$. Let $\upsilon \in (0, 1]$ be such that $\int f^\upsilon\,\mathrm{d}\lambda < \infty$. For any fixed $\sigma > 0$, $\|f*L_\sigma - f\|_p \lesssim e^{-(\rho/\sigma)^r/2}1_{\{\infty\}}(S_f)$ for every $1 \le p \le \infty$.*

PROOF. We have $(f*L_\sigma - f)(x) = (2\pi)^{-1}\int_{|t|>1/\sigma}e^{-itx}\hat{f}(t)[\hat{L}(t)-1]\,\mathrm{d}t$, $x \in \mathbb{R}$. If $S_f < 1/\sigma$, then the integral is identically equal to zero and $\|f*L_\sigma - f\|_p = 0$ for every $1 \le p \le \infty$. If $S_f = \infty$, for every $2 \le p < \infty$, repeat the same reasoning as for the sinc kernel to conclude that $\|f*L_\sigma - f\|_p^p \le C_p\|\hat{f}(\widehat{L_\sigma} - 1)\|_q^q = C_p\int_{|t|>1/\sigma}(|\hat{f}(t)||\hat{L}(t)-1|)^q\,\mathrm{d}t < C_p\int_{|t|>1/\sigma}|\hat{f}(t)|^q\,\mathrm{d}t \lesssim e^{-(\rho/\sigma)^r/2}$ because $|\hat{L}| < 1$. Now, we consider the remaining cases where $1 \le p < 2$. From Lemma 1 in Devroye [5], page 2040, and condition (2.1),

$\|f * L_\sigma - f\|_1 \le 2(\int f^\upsilon \,\mathrm{d}\lambda)(\pi^{-1}\int_{|t|>1/\sigma}|\hat{f}(t)|\,\mathrm{d}t)^{1-\upsilon} \lesssim e^{-(\rho/\sigma)^r/2}$. For every other L^p-metric, with $1 < p < 2$, use the interpolation inequality $\|f * L_\sigma - f\|_p \le \max\{\|f * L_\sigma - f\|_1,\, \|f * L_\sigma - f\|_2\}$ to conclude that $\|f * L_\sigma - f\|_p \lesssim e^{-(\rho/\sigma)^r/2}$. $\square$

Before proving Theorem 3.1, a remark is in order. If

$$\int_{\mathbb{R}} |\hat{K}(\sigma t)|\,\mathrm{d}t < \infty, \tag{5.2}$$

then $\int_{\mathbb{R}} |\widehat{f_{F,\sigma}}(t)|\,\mathrm{d}t = (2\pi)^{-1}\int_{\mathbb{R}} e^{-itx}|\hat{F}(t)\hat{K}(\sigma t)|\,\mathrm{d}t < \infty$. If K is super-smooth, *i.e.*, $I_{\rho,r}(K) \le 2\pi L$ for some ρ, $r > 0$ and $0 < L < \infty$, then, not only is requirement (5.2) met, cf. (2.2), but $f_{F,\sigma}$ itself is super-smooth with $I_{\rho\sigma,r}(f_{F,\sigma}) \le 2\pi L/\sigma$. Condition (5.2), involving *only* the kernel, allows to recover *any* convolution $f_{F,\sigma}$ by just inverting its Fourier transform.

Proof of Theorem 3.1. We appeal to Theorem 2 of Giné and Nickl [13], page 2891. Choosing $\gamma_n = 1$ for all $n \in \mathbb{N}$, we have $\delta_n = \varepsilon_n(n\varepsilon_n^2)^{(1-1/p)/2}$. For $s_n = E(n\varepsilon_n^2)^{-1/\gamma}$, $E > 0$ being a suitable constant, let $\mathscr{P}_n := \{f_{F,\sigma} : F \in \mathscr{M}(\mathbb{R}),\ \sigma \ge s_n\}$. For every $f_{F,\sigma} \in \mathscr{P}_n$, we have $I_{\rho_n,r}(f_{F,\sigma}) \le 2\pi L_n$, with $\rho_n := \rho s_n$ and $L_n := L/s_n$. Condition 1(a), *ibidem*, page 2890, for the convolution kernel case is verified for the sinc kernel. In fact, $\mathrm{sinc} \in L^\infty(\mathbb{R})$ since $\|\mathrm{sinc}\|_\infty = 1/\pi < \infty$. Also, $\mathrm{sinc} \in L^2(\mathbb{R})$ because $\int_{\mathbb{R}} \mathrm{sinc}^2(x)\,\mathrm{d}x = 1/\pi < \infty$. Besides, the sinc kernel is continuous and, as shown in Lemma A.4, is of bounded quadratic variation. Let $\mathrm{sinc}_j(f) := f * \mathrm{sinc}_{2^{-j}}$. By Lemma 5.1, for every density $f_{F,\sigma} \in \mathscr{P}_n$ for which $S_{f_{F,\sigma}} < \infty$, whatever sequence $J_n \to \infty$, for n large enough so that $2^{J_n} > S_{f_{F,\sigma}}$, we have $\|\mathrm{sinc}_{J_n}(f_{F,\sigma}) - f_{F,\sigma}\|_p = 0$ for every $1 \le p \le \infty$. For every density $f_{F,\sigma} \in \mathscr{P}_n$ for which $S_{f_{F,\sigma}} = \infty$, taking J_n such that $2^{J_n} = cn\varepsilon_n^2$, with $c \ge 2^{1/r}/(\rho E)$, and using the constraint on γ, we have $\|\mathrm{sinc}_{J_n}(f_{F,\sigma}) - f_{F,\sigma}\|_p \lesssim \exp\{-(\rho s_n 2^{J_n})^r/2\} \lesssim n^{-1} \lesssim \delta_n$. Consequently, for n large enough, $\mathscr{P}_n \subseteq \{f_{F,\sigma} : \|\mathrm{sinc}_{J_n}(f_{F,\sigma}) - f_{F,\sigma}\|_p \le C(K)\delta_n\}$, where $C(K) > 0$ is an appropriate constant depending only on the operator (sinc) kernel. For $E < [(C+4)/d]^{-1/\gamma}$, where $C > 0$ is the constant arising in the small ball probability estimate, we have $(\Pi \otimes G)(\mathscr{P}_n^c) \lesssim e^{-ds_n^{-\gamma}} \lesssim \exp\{-(C+4)n\varepsilon_n^2\}$ and Assumption (1), *ibidem*, page 2891, is fulfilled. $\square$

Proof of Corollary 3.1. Under the stated conditions, Theorem 3.1 holds for G being a point mass at 1 and $f_0 = f_{F_0,1}$. Thus, for every $2 \le p \le \infty$, there exists a sufficiently large constant $0 < M' < \infty$ so that $\Pi(\{F : \|f_{F,1} - f_0\|_p < M'\delta_n\}|X^{(n)}) \to 1$ in P_0^n-probability, where

$\delta_n = \varepsilon_n(n\varepsilon_n^2)^{(1-1/p)/2}$. Since the kernel K satisfies (3.2), by Theorem 2 of Nguyen [25], page 8, for any F such that $\|f_{F,1} - f_0\|_p < M'\delta_n$,

$$W_1(F, F_0) \le W_2(F, F_0) \lesssim (-\log \|f_{F,1} - f_0\|_1)^{-1/r} \lesssim (\log n)^{-1/r},$$

where the second inequality descends from Lemma A.5 applied to $\|f_{F,1} - f_0\|_1$. In fact, for $u > 0$ such that $\mathrm{E}_K[|X|^u] < \infty$, the absolute moment of order u of X under $f_{F,1}$ is finite for *every* $F \in \mathscr{M}(\Theta)$: in fact, $\mathrm{E}_{f_{F,1}}[|X|^u] \le (1 \vee 2^{u-1})\{\mathrm{E}_K[|X|^u] + \int_\Theta |\theta|^u \,\mathrm{d}F(\theta)\} < \infty$, the integral being finite because F is compactly supported on Θ. Hence, for a suitable constant $0 < M < \infty$, the inclusion $\{F : \|f_{F,1} - f_0\|_p < M'\delta_n\} \subseteq \{F : W_1(F, F_0) < M(\log n)^{-1/r}\}$ holds and the assertion follows. ☐

6. Prior estimates. As mentioned in the introduction, estimates of the probability of an L^1-ball, under different priors, are essential to evaluate the prior mass of Kullback-Leibler type balls as in condition (3.1). While for the NI-G process, the expression of the finite-dimensional distributions can be used to estimate the probability of an L^1-ball along the lines of Lemma A.1 in Ghosal *et al.* [9], pages 518–519, which deals with the Dirichlet process, for the Pitman-Yor process, we exploit the stick-breaking representation to obtain lower bounds on the probabilities of L^1-balls of the mixing weights and the locations, as given in the following two lemmas.

6.1. *Pitman-Yor process.*

LEMMA 6.1. *Let $F \sim \mathrm{PY}(c, d, \bar{\alpha})$, with $c > -d$ and $0 < d < 1$. Let $F' = \sum_{j=1}^N p_j \delta_{z_j}$, $N \ge 1$, be a finite probability measure on $\mathbb{R}$, with $p_1 \ge p_2 \ge \ldots \ge p_N > 0$. Define $v_1 := p_1$ and $v_j := p_j[\prod_{h=1}^{j-1}(1 - v_h)]^{-1}$, $2 \le j \le N$. For $0 < \varepsilon < 1$, let $U := (\sum_{j=1}^N \sum_{h=1}^j |V_h - v_h| \le 2\varepsilon,\ \min_{1\le j\le N} V_j > \varepsilon/N^2)$, where the random variables $V_1, \ldots, V_N$ are those arising from the stick-breaking representation (4.3). There exist constants $c_1, C > 0$ (depending only on c and d) such that, for $(2\varepsilon/N^2) < (1 - p_1)/2$,*

$$\mathrm{P}(U) \ge C \exp\{-c_1 N \max\{\log(N/\varepsilon),\, dN \log(1/(1 - p_1))\}\}.$$

PROOF. If $|V_j - v_j| \le 2\varepsilon/N^2$ for every $j = 1, \ldots, N$, then $\sum_{j=1}^N \sum_{h=1}^j |V_h - v_h| \le 2\varepsilon$. Thus, U is implied by the event $V := (|V_j - v_j| \le 2\varepsilon/N^2,\ V_j > \varepsilon/N^2,\ j = 1, \ldots, N)$. Let $l_j := ((v_j - 2\varepsilon/N^2) \vee (\varepsilon/N^2))$ and $u_j := ((v_j + 2\varepsilon/N^2) \wedge 1)$ for $j = 1, \ldots, N$. By assumption, $V_j \overset{\text{indep}}{\sim} \mathrm{Beta}(1 - d, c + dj)$

for every $j \in \mathbb{N}$, thus, by the identity $\Gamma(z+1) = z\Gamma(z)$, $z > 0$,

$$\begin{aligned}
\mathrm{P}(V) &= \prod_{j=1}^{N} \frac{\Gamma(1-d+c+dj)}{\Gamma(1-d)\Gamma(c+dj)} \int_{l_j}^{u_j} v^{-d}(1-v)^{c+dj-1}\,\mathrm{d}v \\
&= [\Gamma(1-d)]^{-N} \frac{\Gamma(c)}{\Gamma(c+dN)} \prod_{j=1}^{N} [c+d(j-1)] \int_{l_j}^{u_j} v^{-d}(1-v)^{c+dj-1}\,\mathrm{d}v \\
&\geq \frac{[\Gamma(1-d)]^{-N}\Gamma(c)c^N}{\Gamma(c+dN)} \prod_{j=1}^{N} \int_{l_j}^{u_j} (1-v)^{c+dj-1}\,\mathrm{d}v.
\end{aligned}$$

If $N \to \infty$ as $\varepsilon \to 0$, using the formula $\Gamma(c+dN) \sim (2\pi)^{1/2} e^{-dN}(dN)^{dN+c-1/2}$, letting $v_{\max} := \max_{1\leq j\leq N} v_j$,

$$\begin{aligned}
\mathrm{P}(V) &\gtrsim \frac{[\Gamma(1-d)]^{-N}\Gamma(c)c^N(\varepsilon/N^2)^N}{\Gamma(c+dN)} \\
&\qquad\qquad \times\ [1-((v_{\max}+2\varepsilon/N^2)\wedge 1)]^{(c-1)N+dN(N+1)/2} \\
&\gtrsim \frac{[\Gamma(1-d)]^{-N}\Gamma(c)c^N(\varepsilon/N^2)^N}{\Gamma(c+dN)} \times [(1-v_{\max})/2]^{(c-1)N+dN(N+1)/2} \\
&\gtrsim \exp\{-c_1 N \max\{\log(N/\varepsilon),\ dN\log(1/(1-v_{\max}))\}\},
\end{aligned}$$

provided $(2\varepsilon/N^2) < (1-v_{\max})/2$, where $v_{\max} \in (0,\,1)$ because of the positivity assumption on the mixing weights. The assertion follows noting that $v_{\max} = v_1 = p_1$. □

REMARK 6.1. Because of the positivity constraint on d, Lemma 6.1 does not cover the case of a Dirichlet process, which can be treated using Lemma 6.1 in Ghosal *et al.* [9], pages 518–519, or Lemma A.1 in Ghosal [7], pages 1278–1279. Letting $d \to 0$, if $N = O((1/\varepsilon)^{\xi})$, for $\xi > 0$, we have $\mathrm{P}(U) \gtrsim \exp\{-c_1 N \log(1/\varepsilon)\}$, which agrees with the prior estimate known for a Dirichlet process.

LEMMA 6.2. *Let $F \sim \mathrm{PY}(c,\,d,\,\bar{\alpha})$, with $c > -d$, $0 \leq d < 1$ and the (unnormalized) base measure $\alpha = \alpha(\mathbb{R})\bar{\alpha}$ satisfying* (A$_2$) *for some constants $b > 0$ and $0 < \delta < \infty$. For $0 < \varepsilon < 1$, let $F' = \sum_{j=1}^{N} p_j \delta_{z_j}$, $N \geq 1$, be a finite probability measure with* $\mathrm{supp}(F') \subseteq [-a,\,a]$, *for $a \to \infty$. Then,*

$$\mathrm{P}\left(\sum_{j=1}^{N} |Z_j - z_j| \leq \varepsilon\right) \gtrsim \exp\{-N[\log(\alpha(\mathbb{R})/(2\varepsilon)) + ba^{\delta}]\}.$$

PROOF. If $|Z_j - z_j| \le \varepsilon$ for every $j = 1, \dots, N$, then $\sum_{j=1}^N |Z_j - z_j| \le \varepsilon$. Because $Z_1, \dots, Z_N$ are i.i.d. $\bar{\alpha}$,

$$\begin{aligned}\mathrm{P}\left(\sum_{j=1}^{N} |Z_j - z_j| \le \varepsilon\right) &\ge \prod_{j=1}^{N} \int_{z_j-\varepsilon}^{z_j+\varepsilon} \frac{\alpha'(z)}{\alpha(\mathbb{R})}\,\mathrm{d}z \\ &\gtrsim \exp\{-N[\log(\alpha(\mathbb{R})/(2\varepsilon)) + ba^{\delta}]\},\end{aligned}$$

the last inequality following from (A_2) and the assumption that $a \to \infty$. $\square$

6.2. *Normalized inverse-Gaussian process.* We prove an analogue of Lemma 6.1 in Ghosal *et al.* [9], pages 518–519, or Lemma A.1 in Ghosal [7], pages 1278–1279. We provide an estimate of the probability of an L^1-ball in $\mathbb{R}^N$ under the N-IG distribution. For $r > 0$, let $\bar{B}(z_0; r) := \{z \in \mathbb{R}^N : \|z - z_0\|_1 \le r\}$ be the L^1-closed ball centered at z_0 with radius r.

LEMMA 6.3. *Let $Z := (Z_1, \dots, Z_N)$ be distributed according to the N-IG distribution with parameters $(\alpha_1, \dots, \alpha_N)$. Let $z_0 := (z_{10}, \dots, z_{N0}) \in \Delta^{N-1}$. For $0 < \varepsilon < 1$, let $U := (Z \in \bar{B}(z_0; 2\varepsilon),\ \min_{1\le j\le N} Z_j > \varepsilon^2/2)$. Assume that $A\varepsilon^b \le \alpha_j \le 1$ for every $1 \le j \le N$ and some constants $A, b > 0$. If $\min_{1\le j\le N} z_{j0} > \varepsilon$, there exist constants $c, C > 0$ (depending only on A, b and $m := \sum_{j=1}^N \alpha_j$) such that, for $\varepsilon \le 1/N$ and $N \to \infty$ as $\varepsilon \to 0$, $\mathrm{P}(U) \ge C \exp\{-cN \max\{\log(1/\varepsilon), \log(1/(\min_{1\le j\le N} z_{j0} - \varepsilon))\}\}$.*

PROOF. As in the proof of Lemma 6.1 in Ghosal *et al.* [9], pages 518–519, we can assume that $z_{N0} \ge 1/N$. If $|z_j - z_{j0}| \le \varepsilon^2$ for every $j = 1, \dots, N-1$, then $\|z - z_0\|_1 \le 2\varepsilon$ and $z_N \ge \varepsilon^2 > \varepsilon^2/2$. Therefore, U is implied by $V := (|Z_j - z_{j0}| \le \varepsilon^2,\ Z_j > \varepsilon^2/2,\ j = 1, \dots, N-1)$. For $l_j := \big((z_{j0} - \varepsilon^2) \vee (\varepsilon^2/2)\big)$ and $u_j := \big((z_{j0} + \varepsilon^2) \wedge 1\big)$, $j = 1, \dots, N-1$, $\mathrm{P}(V) = \int_{l_1}^{u_1} \cdots \int_{l_{N-1}}^{u_{N-1}} f(z_1, \dots, z_{N-1})\,\mathrm{d}z_1 \cdots \mathrm{d}z_{N-1}$, where $f = \prod_{r=1}^4 h_r$, with the h_r's as in (4.6). Then,

$$\begin{aligned}\mathrm{P}(V) &\ge \frac{e^m (A\varepsilon^b)^N}{2^{N/2-1}\pi^{N/2}} \times (em)^{-N/2} \left(\min_{1\le j\le N} z_{j0} - \varepsilon\right)^{N/2} \times \left(\frac{\varepsilon^2}{2}\right)^{N-1} \\ &\gtrsim \exp\left\{-cN \max\left\{\log(1/\varepsilon), \log\left(1/\big(\min_{1\le j\le N} z_{j0} - \varepsilon\big)\right)\right\}\right\},\end{aligned}$$

where h_1 is bounded below using the constraint $\alpha_j \ge A\varepsilon^b$, while $h_4 \ge 1$ because every $z_j \le 1$, $j = 1, \dots, N$. To bound below h_2, first note that $K_{-N/2}(\cdot) = K_{N/2}(\cdot)$ (see 9.6.6 in Abramowitz and Stegun [1], page 375).

Since, for ε small enough,

$$(\mathcal{A}_N(z_1,\,\ldots,\,z_{N-1}))^{1/2} \le m^{1/2}\left(\min_{1\le j\le N} z_{j0}-\varepsilon\right)^{-1/2} \ll (N/2+1)^{1/2},$$

the approximation $h_2 \sim 2^{N/2-1}\Gamma(N/2)(\mathcal{A}_N(z_1,\,\ldots,\,z_{N-1}))^{-N/4}$ holds (*ibidem*, formula 9.6.9). By Stirling's formula, $h_2 \gtrsim e^{-N/2}m^{-N/4}(\min_{1\le j\le N} z_{j0}-\varepsilon)^{N/4}$. Consequently, $h_2\times h_3 \gtrsim (em)^{-N/2}(\min_{1\le j\le N} z_{j0}-\varepsilon)^{N/2}$. □

7. Proof of Theorem 4.1. We prove the result for the L^1-distance using Theorem 2.1 of Ghosal and van der Vaart [10], page 1239. Next, we deal with the L^p-metrics, for $2\le p\le\infty$, appealing to Theorem 3.1. For the cases where $1<p<2$, the result follows from the interpolation inequality $\|f_{F,\sigma}-f_0\|_p \le \max\{\|f_{F,\sigma}-f_0\|_1,\,\|f_{F,\sigma}-f_0\|_2\} \lesssim n^{-1/2}(\log n)^{\psi}$, with a suitable $\psi>0$.

• L^1-*distance*. We show that conditions (2.8) and (2.9) in Theorem 2.1 of Ghosal and van der Vaart [10], page 1239, are satisfied for $\bar\varepsilon_n = n^{-1/2}(\log n)^{\chi}$, with a suitable χ, and $\tilde\varepsilon_n = n^{-1/2}(\log n)^{\tau+(\tau-1/2)1_{(0,\,\infty)}(d)}$, with τ as in (4.5). Since $\chi>\tau$, the posterior rate is $\varepsilon_n := (\bar\varepsilon_n \vee \tilde\varepsilon_n) = \bar\varepsilon_n$. Given $\eta_n\in(0,\,1/5)$, for constants $E,\,F,\,L>0$ to be suitably chosen, let $s_n = E(\log(1/\eta_n))^{-2\tau/\gamma}$, $S_n = \exp\{F(\log(1/\eta_n))^{2\tau}\}$ and $a_n = L(\log(1/\eta_n))^{2\tau/\delta}$. Using the same sieve set $\mathscr{F}_n$ as in Theorem 4.1 of Scricciolo [30], pages 285–288, with S_n playing the role of t_n, in virtue of Lemma A.10 and the fact that, for $r>1$, $(a_n/s_n)^{r/(r-1)} \gtrsim \log(1/\eta_n)$,

$$\log D(\eta_n,\,\mathscr{F}_n,\,d_{\mathrm{H}}) \lesssim \left(\frac{a_n}{s_n}\right)^{I_{(0,\,1]}(r)+rI_{(1,\,\infty)}(r)/(r-1)} \times \left(\log\frac{1}{\eta_n}\right)^{1+I_{(0,\,1]}(r)/r}.$$

Taking $\eta_n=\bar\varepsilon_n$, we have $\log D(\bar\varepsilon_n,\,\mathscr{F}_n,\,d_{\mathrm{H}}) \lesssim n\bar\varepsilon_n^2$. As for condition (2.9), by assumption (*ii*) and the fact that $2\tau>1$, for appropriate choices of $E,\,F,\,L$ as functions of the constant c_2 arising from the small ball probability, $(\Pi\otimes G)(\mathscr{F}_n^c) \lesssim e^{-ds_n^{-\gamma}} + S_n^{-\varrho} + e^{-ba_n^{\delta}}/\eta_n^2 \lesssim \exp\{-(c_2+4)n\tilde\varepsilon_n^2\}$ because, by Markov's inequality and the independence of $\{W_j\}_{j\ge1}$ and $\{Z_j\}_{j\ge1}$, $\Pi(\{F:F([-a,\,a]^c)>\eta^2/16\}) < (16/\eta^2)\,\mathrm{E}[\sum_{j=1}^{\infty} W_j 1_{[-a,\,a]^c}(Z_j)] \lesssim \alpha([-a,\,a]^c)/\eta^2 \lesssim e^{-ba^{\delta}}/\eta^2$.

• L^p-*metrics*, $2\le p\le\infty$. We have $\delta_n = \tilde\varepsilon_n(n\tilde\varepsilon_n^2)^{(1-1/p)/2}$. Choose $s_n = E(n\tilde\varepsilon_n^2)^{-1/\gamma}$, with $E<[(c_2+4)/d]^{-1/\gamma}$, $c_2>0$ being the constant arising from condition (3.1) and $d>0$ (in this occurrence) the constant appearing in (A_3). Since $f_0 = f_{F_0,\,\sigma_0}$, we have $\|f_0\|_p<\infty$ and, for n large enough, $\|f_0 * \mathrm{sinc}_{2^{-J_n}} - f_0\|_p = O(\delta_n)$ because $f_0\in\mathscr{P}_n$.

• *Small ball probability.* Next, we show that, for $0 < \varepsilon \le [(1/4) \wedge (\sigma_0/2)]$, there exist constants c_1, $c_2 > 0$ so that

$$(\Pi \otimes G)(B_{\mathrm{KL}}(f_0;\, \varepsilon^2)) \ge c_1 \exp\{-c_2(\log(1/\varepsilon))^{2[\tau+(\tau-1/2)1_{(0,\,\infty)}(d)]}\}.$$

A preliminary remark is in order. The case where $\varpi = \infty$ corresponds to F_0 having compact support, *i.e.*, $F_0([-a_0,\, a_0]) = 1$ for some $0 < a_0 < \infty$. Let $a_\varepsilon := a_0^{I_{\{\infty\}}(\varpi)}(c_0^{-1}\log(1/\varepsilon))^{1/\varpi}$ and let F_0^* be the re-normalized restriction of F_0 to $[-a_\varepsilon,\, a_\varepsilon]$. By Lemma A.3 of Ghosal and van der Vaart [10], page 1261, and (4.1), $\|f_{F_0^*,\sigma_0} - f_0\|_1 \lesssim \varepsilon$. We show that there exists a discrete probability measure F_0' on $[-a_\varepsilon,\, a_\varepsilon]$, with at most

$$N \lesssim \left(\log \frac{1}{\varepsilon}\right)^{2\tau-1} \tag{7.1}$$

support points, such that $\|f_{F_0^*,\sigma_0} - f_{F_0',\sigma_0}\|_\infty \lesssim \varepsilon$. The support points of F_0' can be taken to be at least 2ε-separated. We distinguish the case where $0 < r \le 1$ from the case where $r > 1$. In the latter case, the assertion follows immediately from Lemma A.6: in fact, a_ε can be taken to be large enough to satisfy the requirement $(a_\varepsilon/\sigma_0) \ge 1$. If $0 < r \le 1$, Lemma A.6 cannot be directly applied because the requirement on (a_ε/σ_0) may not be met. Yet, an argument similar to the one used in Lemma 2 of Ghosal and van der Vaart [12], page 705, can be adopted. Consider a partition of $[-a_\varepsilon,\, a_\varepsilon]$ into $k = \lceil a_0^{I_{\{\infty\}}(\varpi)}(c_0^{(1-I_{\{\infty\}}(\varpi))/\varpi}\sigma_0)^{-1}(\log(1/\varepsilon))^{1/r-1+I_{(0,\,\infty)}(\varpi)/\varpi}\rceil$ subintervals $I_1, \ldots, I_k$ of equal length $0 < l \le 2\sigma_0(\log(1/\varepsilon))^{-(1-r)/r}$ and, possibly, a final interval I_{k+1} of length $0 \le l_{k+1} < l$. Let J be the total number of intervals in the partition, which can be either k or $k+1$. Write $F_0^* = \sum_{j=1}^J F_0^*(I_j)F_{0,j}^*$, where $F_{0,j}^*$ denotes the re-normalized restriction of F_0^* to I_j. Then, $f_{F_0^*,\sigma_0}(x) = \sum_{j=1}^J F_0^*(I_j) f_{F_{0,j}^*,\sigma_0}(x) = \sum_{j=1}^J F_0^*(I_j)(F_{0,j}^* * K_{\sigma_0})(x)$, $x \in \mathbb{R}$. For every $j = 1, \ldots, J$, by Lemma A.6 (and Remark A.2) applied to every $f_{F_{0,j}^*,\sigma_0}$, with $(a/\sigma) = (l/2)/\sigma_0 \propto (\log(1/\varepsilon))^{-(1-r)/r}$, there exists a discrete distribution $F_{0,j}'$, with at most $N_j \lesssim \log(1/\varepsilon)$ support points, such that $\|f_{F_{0,j}^*,\sigma_0} - f_{F_{0,j}',\sigma_0}\|_\infty \lesssim \varepsilon$. Defined $F_0' := \sum_{j=1}^J F_0^*(I_j)F_{0,j}'$, we have $\|f_{F_0^*,\sigma_0} - f_{F_0',\sigma_0}\|_\infty \le \sum_{j=1}^J F_0^*(I_j)\|f_{F_{0,j}^*,\sigma_0} - f_{F_{0,j}',\sigma_0}\|_\infty \lesssim \varepsilon$, where F_0' has at most $N \lesssim (J \times N_j) \lesssim k \times \log(1/\varepsilon) \lesssim (\log(1/\varepsilon))^{1/r+I_{(0,\,\infty)}(\varpi)/\varpi}$ support points. Combining the result on the total number N of support points of F_0' in the case where $0 < r \le 1$ with the one in the case where $r > 1$, we obtain the bound in (7.1). Let $q > 0$ be such that $\mathrm{E}_K[|X|^q] < \infty$. For any υ such that $(1+q)^{-1} < \upsilon < 1$, by Hölder's inequality, $\int f_{F_0^*,\sigma_0}^\upsilon\,\mathrm{d}\lambda \lesssim (1 +$

$\int_{\mathbb{R}} |x|^q f_{F_0^*,\,\sigma_0}(x)\,\mathrm{d}x)^\upsilon \lesssim \{(1 \vee 2^{q-1})[\sigma_0^q\,\mathrm{E}_K[|X|^q] + \int_{-a_\varepsilon}^{a_\varepsilon} |\theta|^q\,\mathrm{d}F_0^*(\theta)]\}^\upsilon \lesssim a_\varepsilon^{\upsilon q}$, this implying that $\|f_{F_0^*,\,\sigma_0} - f_{F_0',\,\sigma_0}\|_1 \lesssim \varepsilon^{1-\upsilon} a_\varepsilon^{\upsilon q}$ in virtue of Lemma A.7.

Next, we distinguish the case where the prior for the mixing distribution is a Pitman-Yor process with $d = 0$ and $c = \alpha(\mathbb{R})$, *i.e.*, a Dirichlet process, from the case where the prior for the mixing distribution is a Pitman-Yor process with $d > 0$.

• *Dirichlet process*. Represented F_0' as $\sum_{j=1}^N p_j \delta_{\theta_j}$, with $|\theta_j - \theta_k| \geq 2\varepsilon$ for all $j \neq k$, and set $U_j := [\theta_j - \varepsilon,\, \theta_j + \varepsilon]$, $j = 1, \ldots, N$, for every probability measure F on $\mathbb{R}$ such that

$$\sum_{j=1}^{N} |F(U_j) - p_j| \leq \varepsilon, \tag{7.2}$$

and every $\sigma > 0$ such that $|\sigma - \sigma_0| \leq \varepsilon$, we have $\|f_{F,\,\sigma} - f_{F_0',\,\sigma_0}\|_1 \lesssim \|K_\sigma - K_{\sigma_0}\|_1 + \varepsilon/(\sigma \wedge \sigma_0) + \sum_{j=1}^N |F(U_j) - p_j| \lesssim \varepsilon$ in virtue of Lemma A.8, Lemma A.9 and (7.2). Thus, $\|f_{F,\,\sigma} - f_{F_0',\,\sigma_0}\|_1 \lesssim \varepsilon$ and $d_{\mathrm{H}}^2(f_{F,\,\sigma},\, f_0) \leq \|f_{F,\,\sigma} - f_{F_0',\,\sigma_0}\|_1 + \|f_{F_0',\,\sigma_0} - f_{F_0^*,\,\sigma_0}\|_1 + \|f_{F_0^*,\,\sigma_0} - f_0\|_1 \lesssim \varepsilon^{1-\upsilon} a_\varepsilon^{\upsilon q}$. In order to appeal to Theorem 5 of Wong and Shen [33], pages 357–358, we show that, for densities in $S_{N,\,\varepsilon} := \{f_{F,\,\sigma} :\, \sum_{j=1}^N |F(U_j) - p_j| \leq \varepsilon,\ |\sigma - \sigma_0| \leq \varepsilon\}$ and a suitably chosen $\varrho \in (0,\, 1]$, we have $M_\varrho^2 := \int_{\{(f_0/f_{F,\,\sigma}) \geq e^{1/\varrho}\}} f_0 (f_0/f_{F,\,\sigma})^\varrho\,\mathrm{d}\lambda = O((1/\varepsilon)^\xi)$, with $0 \leq \xi \leq \kappa/\varpi$. For every F satisfying (7.2), $F([-a_\varepsilon,\, a_\varepsilon]) > 1/2$, thus, by symmetry and monotonicity of K, $f_{F,\,\sigma}(x) \geq \int_{-a_\varepsilon}^{a_\varepsilon} K_\sigma(x - \theta)\,\mathrm{d}F(\theta) > K_\sigma(|x| + a_\varepsilon)/2$, $x \in \mathbb{R}$. By assumption (A_0), $K(a_\varepsilon) \gtrsim \exp\{-c a_\varepsilon^\kappa\}$ for a_ε large enough, so that

$$\int_{|x| \leq a_\varepsilon} \frac{f_0^{1+\varrho}(x)}{K_\sigma^\varrho(|x| + a_\varepsilon)}\,\mathrm{d}x \lesssim K_{\sigma_0}^{-\varrho}(4a_\varepsilon) \int_{|x| \leq a_\varepsilon} f_0(x)\,\mathrm{d}x \lesssim \exp\{\varrho c (4a_\varepsilon/\sigma_0)^\kappa\}$$

because $|\sigma - \sigma_0| \leq \varepsilon \leq \sigma_0/2$ and $\|f_0\|_\infty < \infty$ by (2.2). Also,

$$\int_{|x| > a_\varepsilon} \frac{f_0^{1+\varrho}(x)}{K_\sigma^\varrho(|x| + a_\varepsilon)}\,\mathrm{d}x \lesssim \int_{|x| > a_\varepsilon} K_{\sigma_0}^{-\varrho}(4|x|)[K_{\sigma_0}(|x|/2) + F_0(\{\theta :\, |\theta| > |x|/2\})]\,\mathrm{d}x < \infty,$$

where the last integral is finite for a suitable choice of ϱ and in virtue of the tail condition (4.1) on F_0 postulated in (i). Thus, the inclusion $S_{N,\,\varepsilon} \subseteq B_{\mathrm{KL}}(f_0;\, c_1 \varepsilon^{1-\upsilon} a_\varepsilon^{\upsilon q} (\log(1/\varepsilon))^2)$ holds. To apply Lemma A.2 of Ghosal and van der Vaart [10], pages 1260–1261, note that, for each $|\theta_j| \leq a_\varepsilon$, by condition (4.2) prescribed in (ii), $\alpha(U_j) \gtrsim \varepsilon e^{-b a_\varepsilon^\delta} \gtrsim \varepsilon^{b'}$ for some constant $b' > 0$

because, when $\varpi < \infty$, we have $0 < \delta \le \varpi$ by assumption. Thus, (3.1) is satisfied for $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^\tau$.

• *Pitman-Yor process with* $d > 0$. We need to modify the arguments to control $\|f_{F,\sigma} - f_{F_0',\sigma_0}\|_1$. To the aim, the stick-breaking representation of F is exploited. Let $F_0' = \sum_{j=1}^N p_j\delta_{\theta_j}$ be the finite approximating distribution of F_0^*. By relabelling, we can assume that $p_1 \ge p_2 \ge \ldots \ge p_N \ge 0$. Let $M \le N$ be the number of strictly positive mixing weights p_j. For every $\sigma > 0$, by Lemma A.8 and the inequality $\sum_{j=M+1}^\infty W_j \le \sum_{j=1}^M |W_j - p_j|$,

$$\|f_{F,\sigma} - f_{F_0',\sigma}\|_1 \le 2\sum_{j=1}^M |W_j - p_j| + \frac{2\|K\|_\infty}{\sigma}\sum_{j=1}^M p_j|Z_j - \theta_j|. \tag{7.3}$$

Let $v_1 := p_1$ and $v_j := p_j[\prod_{h=1}^{j-1}(1-v_h)]^{-1}$ for $j = 2, \ldots, M$. Note that $v_j \in (0, 1)$ for every $j = 1, \ldots, M$. By (4.3),

$$|W_j - p_j| \le |V_j - v_j|\prod_{h=1}^{j-1}(1-V_h) + v_j\left|\prod_{h=1}^{j-1}(1-V_h) - \prod_{h=1}^{j-1}(1-v_h)\right| \le \sum_{h=1}^{j}|V_h - v_h|,$$

where the inequality $|\prod_{h=1}^{j-1} y_h - \prod_{h=1}^{j-1} z_h| \le \sum_{h=1}^{j-1}|y_h - z_h|$, valid for complex numbers $y_1, \ldots, y_{j-1}$ and $z_1, \ldots, z_{j-1}$ of modulus at most 1, has been used. If, for $0 < \varepsilon \le \sigma_0/2$,

a) $\sum_{j=1}^M\sum_{h=1}^j |V_h - v_h| \le \varepsilon$,
b) $\sum_{j=1}^M |Z_j - \theta_j| \le \varepsilon$,
c) $|\sigma - \sigma_0| \le \varepsilon$,

then $\|f_{F,\sigma} - f_{F_0',\sigma_0}\|_1 \lesssim \|K_\sigma - K_{\sigma_0}\|_1 + \sum_{j=1}^M\sum_{h=1}^j|V_h - v_h| + \sum_{j=1}^M p_j|Z_j - \theta_j| \lesssim \varepsilon$ by Lemma A.9 and (7.3).

Next, we show that, for $B_\varepsilon = a_\varepsilon$ (or $B_\varepsilon = a_\varepsilon + 1$, the latter case being considered if any support point θ_j of F_0' is equal to $-a_\varepsilon$ and/or a_ε), the events in *a*) and *b*) together imply that, for $0 < \varepsilon \le [(1/4) \wedge (\sigma_0/2)]$, we have $F([-B_\varepsilon, B_\varepsilon]) > 1/2$. This inequality is used when checking that, for a suitably chosen $\varrho \in (0, 1]$, $M_\varrho^2 = O((1/\varepsilon)^\xi)$, with $0 \le \xi \le \kappa/\varpi$, so that Theorem 5 of Wong and Shen [33], pages 357–358, can be invoked. By the event in *b*), for ε small enough, all the Z_j's are in $[-B_\varepsilon, B_\varepsilon]$. Using this fact and the inequality $\sum_{j=1}^M |W_j - p_j| \le \sum_{j=1}^M\sum_{h=1}^j|V_h - v_h|$, the event in *a*) implies that $F([-B_\varepsilon, B_\varepsilon]^c) \le \sum_{j=1}^M\sum_{h=1}^j|V_h - v_h| \le \varepsilon < 1/2$.

We estimate the probabilities of the events in *a*) and *b*). In view of the independence of the random variables W_j's and Z_j's, by Lemma 6.1 and Lemma 6.2, for $(1-p_1) > (4\varepsilon/N^2)$ (in case p_1 does not satisfy the condition,

$f_{F'_0,\sigma}$ can be projected into a new density $f_{F''_0,\sigma}$ which is within L^1-distance ε from $f_{F,\sigma}$. This new density can be obtained by first changing the point mass p_1 to p'_1 such that $(1-p'_1) > (4\varepsilon/N^2)$ and then equally distributing the exceeding mass among the other $M-1$ points),

$$\begin{aligned}
&\mathrm{P}\left(\sum_{j=1}^{M}\sum_{h=1}^{j}|V_h - v_h| \le \varepsilon\right) \times \mathrm{P}\left(\sum_{j=1}^{M}|Z_j - \theta_j| \le \varepsilon\right) \\
&\qquad \gtrsim \exp\{-c_1 M^2 \log(1/(1-p_1))\} \times \exp\{-M[\log(\alpha(\mathbb{R})/(2\varepsilon)) + b a_\varepsilon^\delta]\} \\
&\qquad \gtrsim \exp\{-c_2 M^2 \log(1/\varepsilon)\},
\end{aligned}$$

because, by (7.1), $M \le N \lesssim (\log(1/\varepsilon))^{2\tau-1}$, where $\tau \ge 1$, and, for $\varpi < \infty$, we have $0 < \delta \le \varpi$ by assumption, so that $a_\varepsilon^\delta \lesssim \log(1/\varepsilon)$. Thus, $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^{2\tau-1/2}$. □

8. Proof of Theorem 4.3. The main difficulty in the proof rests in finding a finite mixing distribution, with a sufficiently restricted number of support points, such that the corresponding Gaussian mixture approximates the true density, in Kullback-Leibler divergence, with an exponentially small error. Such a mixing distribution may be found by matching the moments of an *ad hoc* constructed mixing density. The crux is the approximation of an analytic density with exponentially decaying Fourier transform by convolving the Gaussian kernel with an operator, whose expression resembles a Taylor series expansion, with suitably calibrated coefficients and derivatives convolved with the sinc kernel. Such a (not necessarily non-negative) function turns out to be a convex linear combination of iterated convolutions of the true density with the Gaussian kernel, which has the effect of reducing the bias. Once this function is modified to be a density with sub-Gaussian tails, the compactly supported version obtained by re-normalizing the restriction to a compact set is employed to find an approximating mixing distribution.

We start by stating the auxiliary result on the approximation of analytic densities by convolutions with the Gaussian kernel. Let $m_j := \int_{\mathbb{R}} y^j \phi(y)\,\mathrm{d}y$ denote the moment of order j of a standard normal. For every $j \in \mathbb{N}$, we define two collections of numbers c_j and d_j as follows. For $j = 1$, we set $c_1 = d_1 = 0$. For $j = 2$, we set $c_2 = 0$ and $d_2 = m_2/2!$. For every integer $j \ge 3$,

$$c_j := -\sum_{\substack{j=k+l\\ k\ge 1,\, l\ge 1}} \frac{m_k m_l}{k!l!}, \qquad d_j := \frac{(-1)^j m_j}{j!} + c_j.$$

Since moments of all odd orders are null for the Gaussian kernel,

$$\forall\, j \in \mathbb{N}, \qquad c_{2j} = -\sum_{\substack{2j=2k+2l\\ k\geq 1,\, l\geq 1}} \frac{m_{2k} m_{2l}}{(2k)!(2l)!}, \qquad d_{2j} = \frac{m_{2j}}{(2j)!} + c_{2j}.$$

Note that the numbers c_j and d_j only depend on the moments of ϕ. For $\sigma > 0$ and a function $f \in \mathcal{C}^\infty(\mathbb{R})$, we define the transform

$$T_\sigma(f) := f - \sum_{j=1}^{\infty} d_j \sigma^j (f^{(j)} * \mathrm{sinc}_\sigma).$$

LEMMA 8.1. *For $\sigma > 0$ small enough and every probability density $f \in C^\omega(\mathbb{R})$ with characteristic function satisfying (2.1) for some constants ρ, $r > 0$ and $0 < L < \infty$,*

$$\|T_\sigma(f) * \phi_\sigma - f\|_\infty \lesssim e^{-(\rho/\sigma)^r/2} 1_{\{\infty\}}(S_f). \tag{8.1}$$

Furthermore,

$$T_\sigma(f) = 3f - 3(f * \phi_\sigma) + f * \phi_\sigma * \phi_\sigma + O(e^{-(\rho/\sigma)^r/2} 1_{\{\infty\}}(S_f)). \tag{8.2}$$

Consequently, $\int T_\sigma(f)\,\mathrm{d}\lambda = 1 + o(e^{-(\rho/\sigma)^r/2} 1_{\{\infty\}}(S_f))$.

PROOF. By definition of $T_\sigma(f)$, Taylor's formula and the assumption that $f \in C^\omega(\mathbb{R})$, for every $x \in \mathbb{R}$,

$$\begin{aligned}
&(T_\sigma(f) * \phi_\sigma - f)(x)\\
&= \int_{\mathbb{R}} \left[f(x-y) - f(x) - \sum_{j=1}^{\infty} d_j \sigma^j (f^{(j)} * \mathrm{sinc}_\sigma)(x-y) \right] \phi_\sigma(y)\,\mathrm{d}y\\
&= \sum_{j=1}^{\infty} \left(\frac{(-1)^j m_j}{j!} \sigma^j f^{(j)}(x) - d_j \sigma^j (f^{(j)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x) \right)\\
&= \sum_{j=1}^{\infty} \left(\frac{(-1)^j m_j}{j!} \sigma^j (f^{(j)} - f^{(j)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x) - c_j \sigma^j (f^{(j)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x) \right),
\end{aligned}$$

where, in the last line, the definition of the d_j's is used. For every $j \in \mathbb{N}$, consider the decomposition

$$\begin{aligned}
&(f^{(j)} - f^{(j)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x)\\
&\quad = \frac{1}{2\pi} \int_{|t|>1/\sigma} (-it)^j e^{-itx} \hat{f}(t)\,\mathrm{d}t\\
&\qquad\qquad + \frac{1}{2\pi} \int_{\mathbb{R}} (-it)^j e^{-itx} \hat{f}(t) 1_{[-1,\,1]}(\sigma t)\,\mathrm{d}t - (f^{(j)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x)\\
&\quad =: T_1(j,\, \sigma,\, x) + T_2(j,\, \sigma,\, x).
\end{aligned}$$

By the Cauchy-Schwarz inequality and the assumption that $\hat{f}$ satisfies (2.1), $T_1(j,\,\sigma,\,x) \lesssim \sigma^{-j}e^{-(\rho/\sigma)^r/2}1_{\{\infty\}}(S_f)$ for σ small enough. Thus,

$$\sum_{j=1}^{\infty}\frac{(-1)^j m_j}{j!}\sigma^j T_1(j,\,\sigma,\,x) \lesssim e^{-(\rho/\sigma)^r/2}1_{\{\infty\}}(S_f)$$

because $\sum_{j=1}^{\infty} m_j/j! < \infty$. Next, we show that $\sum_{j=1}^{\infty}[(-1)^j m_j\sigma^j T_2(j,\,\sigma,\,x)/j! - c_j\sigma^j(f^{(j)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x)] = 0$ identically. Algebra leads to $T_2(j,\,\sigma,\,x) = -\sum_{k=1}^{\infty} m_{2k}\sigma^{2k}(f^{(j+2k)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x)/(2k)!$. Hence,

$$\begin{aligned}
&\sum_{j=1}^{\infty}\frac{(-1)^j m_j}{j!}\sigma^j T_2(j,\,\sigma,\,x)\\
&\qquad\qquad = -\sum_{j=1}^{\infty}\frac{m_{2j}}{(2j)!}\left(\sum_{k=1}^{\infty}\frac{m_{2k}}{(2k)!}\sigma^{2(j+k)}(f^{(2j+2k)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x)\right)\\
&\qquad\qquad = \sum_{s=2}^{\infty}\left(-\sum_{2k+2l=2s}\frac{m_{2k}}{(2k)!}\frac{m_{2l}}{(2l)!}\right)\sigma^{2s}(f^{(2s)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x)\\
&\qquad\qquad = \sum_{s=2}^{\infty} c_{2s}\sigma^{2s}(f^{(2s)} * \mathrm{sinc}_\sigma * \phi_\sigma)(x)
\end{aligned}$$

by definition of the numbers c_{2s}. The proof of (8.1) is thus complete.

Next, we prove (8.2). Because $T_1(j,\,\sigma,\,x) \lesssim \sigma^{-j}e^{-(\rho/\sigma)^r/2}1_{\{\infty\}}(S_f)$ for σ small enough, $T_\sigma(f) = f - \sum_{j=1}^{\infty} d_j\sigma^j f^{(j)} + O(e^{-(\rho/\sigma)^r/2}1_{\{\infty\}}(S_f))$. By definition of the d_j's and taking into account that $\sum_{j=1}^{\infty}(-1)^j m_j\sigma^j f^{(j)}/j! = f * \phi_\sigma - f$, we have $f - \sum_{j=1}^{\infty} d_j\sigma^j f^{(j)} = f - (f * \phi_\sigma - f) - \sum_{j=1}^{\infty} c_j\sigma^j f^{(j)} = 2f - f * \phi_\sigma - \sum_{j=1}^{\infty} c_j\sigma^j f^{(j)}$, where

$$\begin{aligned}
\sum_{j=2}^{\infty} c_{2j}\sigma^{2j}f^{(2j)} &= -\sum_{j=1}^{\infty}\frac{m_{2j}}{(2j)!}\sigma^{2j}\left(\sum_{k=1}^{\infty}\frac{m_{2k}}{(2k)!}\sigma^{2k}f^{(2j+2k)}\right)\\
&= -\sum_{j=1}^{\infty}\frac{m_{2j}}{(2j)!}\sigma^{2j}(f^{(2j)} * \phi_\sigma - f^{(2j)})\\
&= -(f * \phi_\sigma - f) * \phi_\sigma + (f * \phi_\sigma - f)\\
&= -f * \phi_\sigma * \phi_\sigma + 2(f * \phi_\sigma) - f.
\end{aligned}$$

Relationship (8.2) follows. To bound above $\int T_\sigma(f)\,\mathrm{d}\lambda$, note that the coefficients in (8.2) sum up to 1. Also, for σ small enough,

$$\int_{\mathbb{R}} T_1(j,\,\sigma,\,x)1_{\{z:\,T_1(j,\,\sigma,\,z)\neq 0\}}(x)\,\mathrm{d}x = o(\sigma^{-j}e^{-(\rho/\sigma)^r/2})$$

because $\lim_{\sigma\to 0} T_1(j,\, \sigma,\, x) = 0$ identically so that $\lambda(\{z:\, T_1(j,\, \sigma,\, z) \neq 0\}) = o(1)$. $\square$

REMARK 8.1. The technique developed in Lemma 8.1, properly combined with the arguments of Lemma 3.4 of de Jonge and van Zanten [4], pages 3317–3318, can be exploited to find an approximation of a density f contained in a Sobolev space of order β, using the fact that the Fourier transform satisfies $\int_{\mathbb{R}}(1+|t|^2)^\beta |\hat{f}(t)|^2\,\mathrm{d}t < \infty$. We do not report on this aspect in the present article.

Let f_0 be a probability density satisfying (a). Given $\alpha,\, \beta \in (0,\, 1)$ and $\sigma > 0$, we define the sets

$$\begin{aligned} B_\sigma &:= \{x \in \mathbb{R}:\ f_0(x) \geq B\sigma^{-M} e^{-c_1(1/\sigma)^{r_0}}\}, \qquad \text{with } 0 < c_1 \leq \rho_0^{r_0}/2,\\ G_\sigma &:= \{x \in \mathbb{R}:\ T_\sigma(f_0)(x) > \alpha f_0(x)\},\\ U_\sigma &:= \{x \in \mathbb{R}:\ |f_0^{(j)}(x)| \leq D(j,\, \sigma),\, j \in \mathbb{N},\ Q(x) \leq \sigma^{-\beta}\}, \end{aligned}$$

where $D(j,\, \sigma) := \sigma^{-j} e^{c_2 j(1/\sigma)^{r_0}} 2^{-j/r_0} \rho_0^{-j} [\Gamma((2j+1)/r_0)]^{1/2}$. The function $T_\sigma(f_0)$ is modified to be non-negative by setting it equal to a multiple of f_0 when it is below it. Thus, we define $g_\sigma := T_\sigma(f_0) 1_{G_\sigma} + \alpha f_0 1_{G_\sigma^c}$.

LEMMA 8.2. *Suppose f_0 is a probability density satisfying (a)-(b) for $q \geq 2$, $0 < r_0 < q/(3q-2)$ and $r_0(q-1)/q < \beta < 1 - r_0/2$. For fixed $\zeta \in (0,\, 1)$, let $\alpha := C_\zeta^2/2$, where C_ζ is the constant defined in Lemma A.11. Then, for $\sigma > 0$ small enough, $\int g_\sigma\,\mathrm{d}\lambda > \alpha$ and $\int g_\sigma\,\mathrm{d}\lambda = 1 + O(e^{-c_3(1/\sigma)^{r_0}})$ for a suitable constant $c_3 > 0$.*

PROOF. By definition, $g_\sigma > \alpha f_0(1_{G_\sigma} + 1_{G_\sigma^c}) = \alpha f_0$ so that $\int g_\sigma\,\mathrm{d}\lambda > \alpha$. Because g_σ can be written as $g_\sigma = T_\sigma(f_0) + [\alpha f_0 - T_\sigma(f_0)]\, 1_{G_\sigma^c}$, by (8.2),

$$\begin{aligned} \int g_\sigma\,\mathrm{d}\lambda &= 1 + o(e^{-(\rho_0/\sigma)^{r_0}/2} 1_{\{\infty\}}(S_{f_0})) + \int [\alpha f_0 - T_\sigma(f_0)] 1_{G_\sigma^c}\,\mathrm{d}\lambda\\ &= 1 + O(e^{-c_3(1/\sigma)^{r_0}}), \end{aligned}$$

since, for a suitable constant $c > 0$,

$$\int [\alpha f_0 - T_\sigma(f_0)] 1_{G_\sigma^c}\,\mathrm{d}\lambda = O(e^{-c(1/\sigma)^{r_0}}). \tag{8.3}$$

This statement is proved using both the assumption that $\log f_0$ is locally Lipschitz continuous and the monotonicity assumption (b). For a fixed constant $k > 0$ and a fixed $x \in \mathbb{R}$, let $N_{x,\,\sigma} := \{y \in \mathbb{R}:\ |y - x| \leq k\sigma^{1-r_0/2}\}$.

Let σ be small enough so that $k\sigma^{1-r_0/2} \le \delta$. For every $y \in N_{x,\sigma}$, we have $f_0(y) \le f_0(x) \exp\{Q(x)|y-x|\}$. Consequently,

$$(f_0 * \phi_\sigma)(x) \le f_0(x) \int_{N_{x,\sigma}} \exp\{Q(x)|y-x|\}\phi_\sigma(x-y)\,\mathrm{d}y + ((f_0 1_{N^c_{x,\sigma}}) * \phi_\sigma)(x).$$

For $x \in U_\sigma$ and a suitable constant $C > 0$,

$$\begin{aligned}\int_{N_{x,\sigma}} \exp\{Q(x)|y-x|\}\phi_\sigma(x-y)\,\mathrm{d}y &\le \int_{N_{x,\sigma}} [1 + CQ(x)|y-x|]\phi_\sigma(x-y)\,\mathrm{d}y \\ &\le 1 + C\sigma Q(x) \;=\; 1 + O(\sigma^H),\end{aligned}$$

with $H := (1-\beta) > 0$. Also, $((f_0 1_{N^c_{x,\sigma}}) * \phi_\sigma) \lesssim e^{-k_1(1/\sigma)^{r_0}}$ for a suitable constant $k_1 > 0$. Therefore,

$$\forall\, x \in U_\sigma, \qquad (f_0 * \phi_\sigma)(x) \le f_0(x)[1 + O(\sigma^H)] + C_1 e^{-k_1(1/\sigma)^{r_0}}, \tag{8.4}$$

the complement set U^c_σ having exponentially small probability. In fact,

$$\mathrm{P}_0(U^c_\sigma) \le \mathrm{P}_0(\exists\, j \in \mathbb{N}: |f_0^{(j)}(X)| > D(j,\sigma)) + \mathrm{P}_0(Q(X) > \sigma^{-\beta}) =: P_1 + P_2.$$

By (2.1),

$$\forall\, j \in \mathbb{N}, \quad |f_0^{(j)}(x)| \lesssim 2^{-j/r_0}\rho_0^{-j}[\Gamma((2j+1)/r_0)]^{1/2} =: B_0(j), \quad x \in \mathbb{R}, \tag{8.5}$$

which, together with Markov's inequality, implies that

$$P_1 \le \sum_{j\ge1} [D(j,\sigma)]^{-1}\,\mathrm{E}_0[|f_0^{(j)}(X)|] \lesssim \sum_{j\ge1} \sigma^j e^{-c_2 j(1/\sigma)^{r_0}} \lesssim e^{-k_2(1/\sigma)^{r_0}}.$$

Also, $P_2 \le e^{-(1/\sigma)^{\beta q/(q-1)}}\,\mathrm{E}_0[\exp\{[Q(X)]^{q/(q-1)}\}] \lesssim e^{-(1/\sigma)^{r_0}}$. Hence, $\mathrm{P}_0(U^c_\sigma) \lesssim e^{-k_3(1/\sigma)^{r_0}}$. Next, we show that $B_\sigma \cap U_\sigma \subseteq G_\sigma$. Let $\zeta \in (0,1)$ be fixed. For σ small enough and every $x \in B_\sigma \cap U_\sigma$, by (8.2), (8.4) and Lemma A.11,

$$\begin{aligned}T_\sigma(f_0) &= 3f_0 - 3(f_0 * \phi_\sigma) + f_0 * \phi_\sigma * \phi_\sigma + O(e^{-(\rho_0/\sigma)^{r_0}/2} 1_{\{\infty\}}(S_{f_0})) \\ &\ge (C^2_\zeta - 3O(\sigma^H) - 3C_1 e^{-k_1(1/\sigma)^{r_0}} - C_2\sigma^M) f_0 > C^2_\zeta f_0/2,\end{aligned}$$

because, on B_σ, we have $e^{-(\rho_0/\sigma)^{r_0}/2}/f_0 = O(\sigma^M)$. Taking $\alpha = C^2_\zeta/2$, we have $T_\sigma(f_0) > \alpha$ and the inclusion $B_\sigma \cap U_\sigma \subseteq G_\sigma$ holds. To prove (8.3), it suffices to show that $\int [\alpha f_0 - T_\sigma(f_0)] 1_{B^c_\sigma \cup U^c_\sigma}\,\mathrm{d}\lambda \lesssim e^{-c(1/\sigma)^{r_0}}$. This can be shown using the bounds $\mathrm{P}_0(U^c_\sigma) \lesssim e^{-k_3(1/\sigma)^{r_0}}$, $\mathrm{P}_0(B^c_\sigma) \lesssim (\sigma^{-M} e^{-c_1(1/\sigma)^{r_0}})^\gamma$ for every $\gamma \in (0,1)$, and the fact that, up to $O(e^{-(\rho_0/\sigma)^{r_0}/2} 1_{\{\infty\}}(S_{f_0}))$, the transform $T_\sigma(f_0)$ is a linear combination of f_0, $f_0 * \phi_\sigma$ and $f_0 * \phi_\sigma * \phi_\sigma$. We

begin by showing that $\int_{U_\sigma^c}(f_0 * \phi_\sigma)\,\mathrm{d}\lambda \lesssim e^{-(1/\sigma)^{r_0}/2}$. For random variables $Y \sim f_0$ and $Z \sim \mathrm{N}(0,\,1)$,

$$\int_{U_\sigma^c}(f_0*\phi_\sigma)\,\mathrm{d}\lambda \le \mathrm{P}(Y+\sigma Z \in U_\sigma^c,\, |Z| \le \sigma^{-r_0/2})+\mathrm{P}(|Z|>\sigma^{-r_0/2}) =: T_1+T_2,$$

where $T_2 \lesssim e^{-(1/\sigma)^{r_0}/2}$ and $T_1 = 0$ because, for every $x \in U_\sigma^c$, there exists (at least) one $j \in \mathbb{N}$ such that $|f_0^{(j)}(x)| > D(j,\,\sigma) \to \infty$ as $\sigma \to 0$. On the other hand, for every $j \in \mathbb{N}$, for $|Z| \le \sigma^{-r_0/2}$, using the bound $B_0(j)$ in (8.5) with the greatest value $r \ge 2$ that guarantees $\sum_{k=1}^\infty |f_0^{(j+k)}(Y)|/k! < \infty$,

$$|f_0^{(j)}(Y+\sigma Z)| \le |f_0^{(j)}(Y)|+\sigma^{1-r_0/2}\sum_{k=1}^{\infty}\frac{|f_0^{(j+k)}(Y)|}{k!} \lesssim 1+\sigma^{1-r_0/2} \to c < \infty,$$

which implies that $|f_0^{(j)}|$ is bounded away from ∞, thus contradicting the previous statement. By the result just shown, $\int_{B_\sigma^c}(f_0*\phi_\sigma)\,\mathrm{d}\lambda \lesssim \int_{B_\sigma^c\cap U_\sigma}(f_0 * \phi_\sigma)\,\mathrm{d}\lambda + e^{-(1/\sigma)^{r_0}/2}$, where, for $\xi > 1$,

$$\begin{aligned}\int_{B_\sigma^c\cap U_\sigma}(f_0*\phi_\sigma)\,\mathrm{d}\lambda &\le \mathrm{P}(Y+\sigma Z \in B_\sigma^c \cap U_\sigma,\, |Z| \le \sigma^{-r_0/2},\, Y \in B_{\xi\sigma}\cap U_\sigma)\\ &\quad + \mathrm{P}(Y \in U_\sigma^c) + P(Y \in B_{\xi\sigma}^c) + \mathrm{P}(|Z| > \sigma^{-r_0/2})\\ &\lesssim e^{-k_4(1/\sigma)^{r_0}}.\end{aligned}$$

The probability of the first event can be shown to be equal to 0 similarly to Kruijer *et al.* [20], page 1251: on the one hand, since $Y+\sigma Z \in B_\sigma^c$ and $Y \in B_{\xi\sigma}$, we have $|\log f_0(Y+\sigma Z)-\log f_0(Y)| > (1-\xi^{-r_0})\sigma^{-r_0} \to \infty$. On the other hand, since $Y \in U_\sigma$ and $|Z| \le \sigma^{-r_0/2}$, $|\log f_0(Y+\sigma Z) - \log f_0(Y)| \le Q(Y)\sigma|Z| \le \sigma^{(1-r_0/2)-\beta} \to 0$, in contradiction with the previous statement. Analogously, it can be shown that $\int_{B_\sigma^c\cup U_\sigma^c}(f_0 * \phi_\sigma * \phi_\sigma)\,\mathrm{d}\lambda \lesssim e^{-c'(1/\sigma)^{r_0}}$, which completes the proof. ∎

Next, a finite Gaussian mixture is constructed from g_σ such that it approximates the true density, in Kullback-Leibler divergence, with an exponentially small error.

LEMMA 8.3. *Suppose f_0 is a probability density satisfying (a)-(c) for q, r_0 and β as in Lemma 8.2. Then, for $\sigma > 0$ small enough, there exists a finite Gaussian mixture m_σ having at most $N_\sigma = O((a_\sigma/\sigma)^2)$ support points in $[-a_\sigma,\, a_\sigma]$, with $a_\sigma = O(\sigma^{-r_0/2})$, such that, for constants S, $c_5 > 0$,*

$$\max\{\mathrm{KL}(f_0;\, m_\sigma),\, \mathrm{E}_0[(\log(f_0/m_\sigma))^2]\} \lesssim \sigma^{-S}e^{-c_5(1/\sigma)^{r_0}}. \tag{8.6}$$

Proof. We give the proof only for the bound on $\mathrm{KL}(f_0;\, m_\sigma)$, which is decomposed into the sum of three integrals, see (8.8) below. Fix $\zeta \in (0,\,1)$ and let $\alpha := C_\zeta^2/2$. Set $C_{g_\sigma} := \int g_\sigma\,\mathrm{d}\lambda$, by Lemma 8.2, $C_{g_\sigma} = 1 + Ae^{-c_3(1/\sigma)^{r_0}}$, where $A > 0$ is a suitable constant. Defined the probability density $h_\sigma := g_\sigma/C_{g_\sigma}$,

$$\forall\,\sigma < \tau_\zeta, \qquad h_\sigma * \phi_\sigma \geq \frac{\alpha(f_0 * \phi_\sigma)}{1 + Ae^{-c_3(1/\sigma)^{r_0}}} \geq \frac{\alpha C_\zeta}{1 + Ae^{-c_3(1/\sigma)^{r_0}}} f_0,$$

because $g_\sigma > \alpha f_0$ and Lemma A.11 applies. Furthermore, $|h_\sigma * \phi_\sigma - f_0| \leq C_{g_\sigma}^{-1}|g_\sigma * \phi_\sigma - f_0| + |C_{g_\sigma}^{-1} - 1| f_0 \lesssim |g_\sigma * \phi_\sigma - f_0| + e^{-c_3(1/\sigma)^{r_0}} f_0$. Lemma 8.1 and $\int [\alpha f_0 - T_\sigma(f_0)] 1_{G_\sigma^c}\,\mathrm{d}\lambda \leq \int [\alpha f_0 - T_\sigma(f_0)] 1_{B_\sigma^c \cup U_\sigma^c}\,\mathrm{d}\lambda \lesssim e^{-c(1/\sigma)^{r_0}}$ imply

$$\begin{aligned}|g_\sigma * \phi_\sigma - f_0| &\leq |T_\sigma(f_0) * \phi_\sigma - f_0| + |[(\alpha f_0 - T_\sigma(f_0)) 1_{G_\sigma^c}] * \phi_\sigma| \\ &\lesssim e^{-(\rho_0/\sigma)^{r_0}/2} 1_{\{\infty\}}(S_{f_0}) + \sigma^{-1} e^{-c(1/\sigma)^{r_0}}.\end{aligned}$$

Therefore,

$$\|h_\sigma * \phi_\sigma - f_0\|_\infty \lesssim e^{-c_4(1/\sigma)^{r_0}}. \tag{8.7}$$

Now, consider

$$\mathrm{KL}(f_0;\, h_\sigma * \phi_\sigma) = \left(\int_{B_\sigma \cap U_\sigma} + \int_{B_\sigma^c \cup U_\sigma^c}\right)\left(f_0 \log \frac{f_0}{h_\sigma * \phi_\sigma}\right)\mathrm{d}\lambda =: I_1 + I_2.$$

For $c_1 < c_4$, by (8.7),

$$\begin{aligned}I_1 &\leq \frac{\sup_{x \in B_\sigma \cap U_\sigma} |f_0(x) - (h_\sigma * \phi_\sigma)(x)|}{\inf_{x \in B_\sigma} f_0(x) - \sup_{x \in B_\sigma \cap U_\sigma} |f_0(x) - (h_\sigma * \phi_\sigma)(x)|} \int_{B_\sigma \cap U_\sigma} f_0\,\mathrm{d}\lambda \\ &\lesssim \frac{e^{-c_4(1/\sigma)^{r_0}}}{e^{-c_1(1/\sigma)^{r_0}}(B\sigma^{-M} - De^{-(c_4 - c_1)(1/\sigma)^{r_0}})} \lesssim e^{-(c_4 - c_1)(1/\sigma)^{r_0}}.\end{aligned}$$

Because $\int_{B_\sigma^c \cup U_\sigma^c} f_0\,\mathrm{d}\lambda \lesssim \sigma^{-\gamma M} e^{-\gamma c_1(1/\sigma)^{r_0}} + e^{-k_3(1/\sigma)^{r_0}}$,

$$I_2 \lesssim (\sigma^{-\gamma M} e^{-\gamma c_1(1/\sigma)^{r_0}} + e^{-k_3(1/\sigma)^{r_0}}) \log((1 + Ae^{-c_3(1/\sigma)^{r_0}})/(\alpha C_\zeta)),$$

where the logarithmic term is positive because $0 < \alpha C_\zeta < 1$. Thus,

$$\mathrm{KL}(f_0;\, h_\sigma * \phi_\sigma) \lesssim \sigma^{-\gamma M} e^{-\min\{(c_4 - c_1),\, \gamma c_1,\, k_3\}(1/\sigma)^{r_0}}.$$

Next, let $C_{h_\sigma} := \int_{-a_\sigma}^{a_\sigma} h_\sigma\,\mathrm{d}\lambda$ and define $\tilde{h}_\sigma := h_\sigma 1_{[-a_\sigma,\, a_\sigma]}/C_{h_\sigma}$ to be the re-normalized restriction of h_σ to $[-a_\sigma,\, a_\sigma]$. By Lemma A.6, there exists a

discrete distribution $\tilde{F}$ on $[-a_\sigma,\, a_\sigma]$, with at most $N_\sigma = O((a_\sigma/\sigma)^2)$ support points, such that $\|\tilde{h}_\sigma * \phi_\sigma - \tilde{F} * \phi_\sigma\|_\infty \lesssim \sigma^{-1}e^{-N_\sigma}$. Set $\widetilde{m}_\sigma := C_{h_\sigma}(\tilde{F} * \phi_\sigma)$,

$$\begin{aligned}|h_\sigma * \phi_\sigma - \widetilde{m}_\sigma| &\le \|\tilde{h}_\sigma * \phi_\sigma - \tilde{F} * \phi_\sigma\|_\infty + (h_\sigma 1_{[-a_\sigma,\, a_\sigma]^c}) * \phi_\sigma \\ &\le \sigma^{-1}e^{-N_\sigma} + (h_\sigma 1_{[-a_\sigma,\, a_\sigma]^c}) * \phi_\sigma.\end{aligned}$$

Because f_0 satisfies (c), for fixed $\eta \in (0,\, 1)$ and σ small enough, we have $(h_\sigma 1_{[-a_\sigma,\, a_\sigma]^c}) * \phi_\sigma \lesssim e^{-(\rho_0/\sigma)^{r_0}/2} 1_{\{\infty\}}(S_{f_0}) + \phi(\eta a_\sigma)(1_{[a_\sigma,\, a_\sigma]^c} * \phi_\sigma)$ in virtue of Lemma 8.1 and Lemma 14 in Maugis and Michel [23], page 47. Thus, for a suitable constant $c'' > 0$,

$$\|h_\sigma * \phi_\sigma - \widetilde{m}_\sigma\|_\infty \lesssim \sigma^{-1}e^{-N_\sigma} + e^{-(\rho_0/\sigma)^{r_0}/2} 1_{\{\infty\}}(S_{f_0}) + \phi(\eta a_\sigma) \lesssim e^{-c''(1/\sigma)^{r_0}}.$$

Now, define $t := \widetilde{m}_\sigma + D_\sigma \phi_\sigma$, with $D_\sigma := \sigma^{-(R-1)} e^{-\tilde{c}(1/\sigma)^{r_0}}$, $R > 1$, and the finite Gaussian mixture with density

$$m_\sigma := \frac{t}{\int t \,\mathrm{d}\lambda} = \frac{\widetilde{m}_\sigma + D_\sigma \phi_\sigma}{C_{h_\sigma} + D_\sigma}.$$

Write

$$\begin{aligned}\mathrm{KL}(f_0;\, m_\sigma) \quad &= \quad \mathrm{KL}(f_0;\, h_\sigma * \phi_\sigma) + \int f_0 \log \frac{h_\sigma * \phi_\sigma}{t} \,\mathrm{d}\lambda + \int f_0 \log \frac{t}{m_\sigma} \,\mathrm{d}\lambda \\ &=: \quad J_1 + J_2 + J_3. \qquad (8.8)\end{aligned}$$

- *Control of* J_1. It has been shown that $J_1 \lesssim \sigma^{-\gamma M} e^{-\min\{(c_4 - c_1),\, \gamma c_1,\, k_3\}(1/\sigma)^{r_0}}$.
- *Control of* J_2. Write $J_2 = (\int_{B_\sigma} + \int_{B_\sigma^c}) f_0 \log((h_\sigma * \phi_\sigma)/t) \,\mathrm{d}\lambda =: J_{21} + J_{22}$,

where, for $c_1 < (c'' \wedge \tilde{c})$,

$$\begin{aligned}J_{21} &\le \int_{B_\sigma} f_0 \frac{h_\sigma * \phi_\sigma - t}{t} \,\mathrm{d}\lambda \\ &\lesssim \frac{\sigma^{-R} e^{-(c''\wedge\tilde{c})(1/\sigma)^{r_0}}}{B\sigma^{-M} e^{-c_1(1/\sigma)^{r_0}} - e^{-c''(1/\sigma)^{r_0}}} \int_{B_\sigma} f_0 \,\mathrm{d}\lambda \\ &\lesssim \sigma^{M-R} e^{-[(c''\wedge\tilde{c}) - c_1](1/\sigma)^{r_0}},\end{aligned}$$

because $|h_\sigma * \phi_\sigma - t| \le |h_\sigma * \phi_\sigma - \widetilde{m}| + D_\sigma \phi_\sigma \lesssim \sigma^{-R} e^{-(c''\wedge\tilde{c})(1/\sigma)^{r_0}}$ and, over B_σ, $h_\sigma * \phi_\sigma \gtrsim f_0 \gtrsim B\sigma^{-M} e^{-c_1(1/\sigma)^{r_0}}$ so that $t > \widetilde{m}_\sigma \ge h_\sigma * \phi_\sigma - |h_\sigma * \phi_\sigma - \widetilde{m}_\sigma| \gtrsim \sigma^{-M} e^{-c_1(1/\sigma)^{r_0}} - e^{-c''(1/\sigma)^{r_0}}$. Because $\|h_\sigma * \phi_\sigma\|_\infty < \infty$ and $t \ge D_\sigma \phi_\sigma$,

$$\begin{aligned}J_{22} &\lesssim \log(\sigma/D_\sigma) \int_{B_\sigma^c} f_0 \,\mathrm{d}\lambda + \frac{1}{2\sigma^2} \int_{B_\sigma^c} x^2 f_0(x) \,\mathrm{d}x \\ &\lesssim \sigma^{-(\gamma M + r_0)} e^{-\gamma c_1 (1/\sigma)^{r_0}} + \sigma^{-(\gamma M + 2)} e^{-\gamma c_1 (1/\sigma)^{r_0}} \\ &\lesssim \sigma^{-[\gamma M + (2\vee r_0)]} e^{-\gamma c_1 (1/\sigma)^{r_0}}.\end{aligned}$$

• *Control of* J_3. Noting that $t/m_\sigma = C_{h_\sigma} + D_\sigma \leq 1 + D_\sigma$, we have $J_3 \leq \log(1 + D_\sigma) \leq D_\sigma = \sigma^{-(R-1)} e^{-\tilde{c}(1/\sigma)^{r_0}}$.

Combining partial results, for $c_1 < \min\{c_4,\, c'',\, \tilde{c}\}$, we have $\mathrm{KL}(f_0;\, m_\sigma) \lesssim \sigma^{-S} e^{-c_5(1/\sigma)^{r_0}}$, where S and c_5 are suitable positive constants. The same reasoning applies to bound $\mathrm{E}_0[(\log(f_0/m_\sigma))^2]$ and (8.6) follows. ☐

Proof of Theorem 4.3. As in Theorem 4.1, we prove the result for the L^1-distance. Next, we deal with the L^p-metrics, for $2 \leq p \leq \infty$. The cases where $1 < p < 2$ are covered by interpolation.

• L^1*-distance*. The entropy condition (2.8) and the remaining mass condition (2.9) in Theorem 2.1 of Ghosal and van der Vaart [10], page 1239, can be shown to be satisfied as in Theorem 4.1 and Theorem 4.2 of Scricciolo [30], pages 285–289, computing the remaining mass as in Theorem 4.1.

• L^p*-metrics,* $2 \leq p \leq \infty$. Same proof as in Theorem 4.1.

• *Small ball probability*. We show that, for constants $c_1,\, c_2 > 0$, $(\Pi \otimes G)(B_{\mathrm{KL}}(f_0;\, \tilde{\varepsilon}_n^2)) \geq c_1 \exp\{-c_2 n \tilde{\varepsilon}_n^2\}$, with $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^{\varsigma}$, for a suitable $\varsigma > 0$ independent of r_0. By Lemma 8.3, for σ small enough, there exists a finite Gaussian mixture m_σ, with $N_\sigma \lesssim (a_\sigma/\sigma)^2$ support points $\theta_1,\, \ldots,\, \theta_{N_\sigma}$ in $[-a_\sigma,\, a_\sigma]$, where $a_\sigma = O(\sigma^{-r_0/2})$, such that (8.6) holds. Let $p_1,\, \ldots,\, p_{N_\sigma}$ denote the mixing weights of m_σ. The inequality in (8.6) holds for any Gaussian mixture $m_{\sigma'}$, with $\sigma' \in [\sigma,\, \sigma + e^{-d_1(1/\sigma)^{r_0}}]$, having support points $\theta_1',\, \ldots,\, \theta_{N_\sigma}'$ such that $\sum_{j=1}^{N_\sigma} |\theta_j' - \theta_j| \leq e^{-d_2(1/\sigma)^{r_0}}$ and mixing weights $p_1',\, \ldots,\, p_{N_\sigma}'$ such that $\sum_{j=1}^{N_\sigma} |p_j' - p_j| \leq e^{-d_3(1/\sigma)^{r_0}}$ for suitable constants $d_1,\, d_2,\, d_3 > 0$. Let $\tilde{B}_\sigma := \{f_0 \geq \zeta_\sigma\}$, with $\zeta_\sigma := B'\sigma^{-S'} e^{-c'(1/\sigma)^{r_0}}$, where $S' := (S-2)/\gamma$, with $(1/2) < \gamma < 1$ and $c' < 3c_5$, the constants S and c_5 being those appearing in (8.6), γ being arbitrarily fixed (note that this γ is different from the one appearing in (A_3)). For any probability measure F on $\mathbb{R}$ and $\sigma' \in [\sigma,\, \sigma + e^{-d_1(1/\sigma)^{r_0}}]$,

$$\mathrm{KL}(f_0;\, f_{F,\sigma'}) \lesssim \sigma^{-S} e^{-c_5(1/\sigma)^{r_0}} + \left(\int_{\tilde{B}_\sigma} + \int_{\tilde{B}_\sigma^c}\right)\left(f_0 \log \frac{m_{\sigma'}}{f_{F,\sigma'}}\right). \tag{8.9}$$

We provide an upper bound on the second integral. For any F such that $F([-a_\sigma,\, a_\sigma]) \geq 1/2$, we have $f_{F,\sigma'}(x) \gtrsim \sigma^{-1} \exp\{-(x^2 + a_\sigma^2)/(\sigma')^2\}$ for all $x \in \mathbb{R}$. From Lemma 8.3, $\|m_{\sigma'}\|_\infty \lesssim \sigma^{-1}$. Also, $\int_{\tilde{B}_\sigma^c} (x/\sigma)^2 f_0(x)\, \mathrm{d}x \lesssim \sigma^{-2}\zeta_\sigma^\gamma$ and $\int_{\tilde{B}_\sigma^c} f_0\, \mathrm{d}\lambda \lesssim \zeta_\sigma^\gamma$. Therefore, for a suitable constant $c'' > 0$,

$$\int_{\tilde{B}_\sigma^c} f_0 \log \frac{m_{\sigma'}}{f_{F,\sigma'}}\, \mathrm{d}\lambda \lesssim \int_{\tilde{B}_\sigma^c} \frac{x^2}{\sigma^2} f_0(x)\, \mathrm{d}x + \left(\frac{a_\sigma}{\sigma}\right)^2 \int_{\tilde{B}_\sigma^c} f_0\, \mathrm{d}\lambda \lesssim e^{-c''(1/\sigma)^{r_0}}.$$

• *Dirichlet process*. Clearly,

$$\int_{\tilde{B}_\sigma} f_0 \log(m_{\sigma'}/f_{F,\sigma'})\,\mathrm{d}\lambda \le \int_{\tilde{B}_\sigma} f_0(\|m_{\sigma'} - f_{F,\sigma'}\|_\infty/f_{F,\sigma'})\,\mathrm{d}\lambda.$$

Using Lemma 5 of Ghosal and van der Vaart [12], page 711, we get $\|m_{\sigma'} - f_{F,\sigma'}\|_\infty \lesssim \sigma^{-2}\max_{1\le j\le N_\sigma}\lambda(U_j) + \sigma^{-1}\sum_{j=1}^{N_\sigma}|F(U_j) - p_j|$, where $U_0, \ldots, U_{N_\sigma}$ is a partition of $\mathbb{R}$, with $U_0 := (\bigcup_{j=1}^{N_\sigma} U_j)^c$ and $U_j \ni \theta_j$ for $j = 1, \ldots, N_\sigma$. Since $2a_\sigma/N_\sigma = 2\sigma^2/a_\sigma \gtrsim \sigma^{-S''}e^{-c''(1/\sigma)^{r_0}}$ for some S'' and c'', the support points of $m_{\sigma'}$ can be taken to be at least $\sigma^{-3(S-2)}e^{-3c_5(1/\sigma)^{r_0}}$-separated. If not, $m_{\sigma'}$ can be projected onto a mixture $m'_{\sigma'}$, with $\sigma^{-3(S-2)}e^{-3c_5(1/\sigma)^{r_0}}$-separated points, such that $\|m_{\sigma'} - m'_{\sigma'}\|_\infty \lesssim \sigma^{-(3S-4)}e^{-3c_5(1/\sigma)^{r_0}}$. Thus, we can find disjoint intervals $U_1, \ldots, U_{N_\sigma}$ so that $U_j \ni \theta_j$ and

$$\sigma^{-3(S-2)}e^{-3c_5(1/\sigma)^{r_0}} \le \lambda(U_j) \le 2\sigma^{-3(S-2)}e^{-3c_5(1/\sigma)^{r_0}}, \qquad j = 1, \ldots, N_\sigma.$$

Let F be such that

$$\sum_{j=1}^{N_\sigma}|F(U_j) - p_j| \le \sigma^{-(3S-5)}e^{-3c_5(1/\sigma)^{r_0}}. \tag{8.10}$$

Then, $\|m_{\sigma'} - f_{F,\sigma'}\|_\infty \lesssim \sigma^{-(3S-4)}e^{-3c_5(1/\sigma)^{r_0}}$ and, on $\tilde{B}_\sigma$, $f_{F,\sigma'} \gtrsim m_{\sigma'} - \sigma^{-(3S-4)}e^{-3c_5(1/\sigma)^{r_0}} \gtrsim \zeta_\sigma$. Therefore, $\int_{\tilde{B}_\sigma} f_0\log(m_{\sigma'}/f_{F,\sigma'})\,\mathrm{d}\lambda \lesssim \sigma^{-S}e^{-\mathfrak{c}_4(1/\sigma)^{r_0}}$. Note that for F satisfying (8.10), $F([-a_\sigma, a_\sigma]) \ge 1/2$. Combining partial results, $\max\{\mathrm{KL}(f_0; f_{F,\sigma'}), \mathrm{E}[(\log(f_0/f_{F,\sigma'}))^2]\} \lesssim \sigma^{-S}e^{-c_6(1/\sigma)^{r_0}}$. To apply Lemma A.2 of Ghosal and van der Vaart [10], pages 1260–1261, to estimate the prior probability of $\{F : \sum_{j=1}^{N_\sigma}|F(U_j) - p_j| \le \sigma^{-(3S-5)}e^{-3c_5(1/\sigma)^{r_0}}\}$, note that, since $0 < \delta \le 2$, $\alpha(U_j) \ge \lambda(U_j)\inf_{|\theta|\le a_\sigma}\alpha'(\theta) \gtrsim \sigma^{-3(S-2)}e^{-(3c_5+b)(1/\sigma)^{r_0}}$. Also, $N_\sigma\sigma^{-(3S-5)}e^{-3c_5(1/\sigma)^{r_0}} \lesssim 1$. Thus,

$$\begin{aligned}
&(\Pi\otimes G)(B_{\mathrm{KL}}(f_0; \sigma^{-S}e^{-c_6(1/\sigma)^{r_0}}))\\
&\qquad\gtrsim \Pr([\sigma, \sigma + e^{-d_1(1/\sigma)^{r_0}}])\\
&\qquad\quad\times \Pr\left(\left\{F : \sum_{j=1}^{N_\sigma}|F(U_j) - p_j| \le \sigma^{-(3S-5)}e^{-3c_5(1/\sigma)^{r_0}}\right\}\right)\\
&\qquad\gtrsim \exp\{-d/\sigma^2 - 2d_1(1/\sigma)^{r_0} - c_7N_\sigma(1/\sigma)^{r_0}\} \gtrsim \exp\{-c'N_\sigma(1/\sigma)^{r_0}\},
\end{aligned}$$

for a suitable constant $c' > 0$. Taking $\sigma = (\log n)^{-1/r_0}$, $r_0 = 1/3$ and $S = 8/3$, we find $(\Pi\otimes G)(B_{\mathrm{KL}}(f_0; \tilde{\varepsilon}_n^2)) \gtrsim \exp\{-c_2n\tilde{\varepsilon}_n^2\}$, for $\tilde{\varepsilon}_n^2 = n^{-1/2}(\log n)^4$.

- *Pitman-Yor process with $d>0$*. It is enough to note that

$$\|m_{\sigma'}-f_{F,\sigma'}\|_\infty \lesssim \frac{1}{\sigma}\sum_{j=1}^{M}|W_j-p_j|+\frac{1}{\sigma^2}\sum_{j=1}^{M}p_j|Z_j-\theta_j|$$

and then proceed estimating the probabilities in a) and b) as in Theorem 4.1. Again, the rate is $\tilde{\varepsilon}_n=n^{-1/2}(\log n)^{\iota}$, for a suitable constant $\iota>0$ independent of r_0.
- *NI-G process*. Same proof as for the Dirichlet process. □

APPENDIX: AUXILIARY RESULTS

This Appendix reports some auxiliary results. Proofs that are an adaptation of those of results known in the literature are omitted.

The following proposition establishes the analyticity of EPD's with shape parameter that is an even integer.

Proposition A.1. *For $\theta\in\mathbb{R}$, $\sigma>0$ and $m\in\mathbb{N}$, let $\widehat{f_{\theta,\sigma,2m}}$ be the characteristic function of an* EPD$(\theta,\sigma,2m)$. *Then, $\widehat{f_{\theta,\sigma,2m}}(t)\le Be^{-ct^2}$, $t\in\mathbb{R}$, where $B,c>0$ are constants depending only on σ and m. The corresponding* EPD *is analytic.*

Proof. The assertion trivially holds for $m=1$, which corresponds to the case of a Gaussian distribution. For $m\ge 2$, let $\Psi_{k,m}:=(2k+1)/(2m)$. From (2.4),

$$\widehat{f_{\theta,\sigma,2m}}(t)=\frac{\sqrt{\pi}e^{it\theta}}{\Gamma(1/(2m))}\sum_{k=0}^{\infty}\frac{\Gamma(\Psi_{k,m})}{\Gamma(m\Psi_{k,m})}\times\frac{[-(\sigma t(2m)^{1/(2m)}/2)^2]^k}{k!},\qquad t\in\mathbb{R},$$

cf. (6) in Pogány [28], page 50. Applying Gauss' multiplication formula, see, *e.g.*, Abramowitz and Stegun [1], page 256,

$$\begin{aligned}\widehat{f_{\theta,\sigma,2m}}(t)&=\frac{\sqrt{\pi}e^{it\theta}}{\Gamma(1/(2m))}\\&\qquad\times\sum_{k=0}^{\infty}\frac{\Gamma(\Psi_{k,m})[-(\sigma t(2m)^{1/(2m)}/2)^2]^k/k!}{(2\pi)^{-(m-1)/2}m^{(m\Psi_{k,m}-1/2)}\prod_{j=0}^{m-1}\Gamma(\Psi_{k,m}+j/m)}\\&\le\frac{(2\pi)^{m/2}}{\sqrt{2}\Gamma(1/(2m))}\sum_{k=0}^{\infty}\underbrace{\left[\prod_{j=1}^{m-1}\Gamma\left(\frac{k}{m}+\frac{2j+1}{2m}\right)\right]^{-1}}_{=:T_{k,m}}\\&\qquad\qquad\times\frac{[-(\sigma t(2m)^{1/(2m)}/2)^2]^k}{k!},\end{aligned}$$

because $m^{-k} \leq 2^{-k} \leq 1$ for all $k = 0, 1, \ldots$. Next, an upper bound on $T_{k,m}$, independent of k, is provided. To the aim, note that the Gamma function is continuous on $(0, \infty)$ and attains the minimum at a point $z^* \in (1, 2)$. Therefore, for every $m \geq 2$, we have $T_{k,m} \leq T_{k^*,m}$ for all $k = 0, 1, \ldots$, where k^* is the smallest positive integer such that $1 < [k^* + j + 1/2]/m < 2$, $j = 1, \ldots, m-1$. Thus,

$$\widehat{f_{\theta,\sigma,2m}}(t) \leq \frac{(2\pi)^{m/2} T_{k^*,m}}{\sqrt{2}\Gamma(1/(2m))} \exp\left(-\sigma^2 t^2 (2m)^{1/m}/4\right) =: Be^{-ct^2}, \qquad t \in \mathbb{R}.$$

Hence, $\widehat{f_{\theta,\sigma,2m}}(t) \lesssim e^{-c|t|}$ for large $|t|$, and, in virtue of Theorem 11.7.1 in Kawata [19], page 439, the corresponding distribution function is analytic. $\square$

In the next lemma, the sinc kernel is shown to have bounded quadratic variation. By definition, a function h is of bounded p-variation on $\mathbb{R}$, $p \geq 1$ real, if $v_p(h) := \sup\{(\sum_{k=1}^n |h(x_k) - h(x_{k-1})|^p)^{1/p} : -\infty < x_0 < \ldots < x_n < \infty,\ n \in \mathbb{N}\}$ is finite.

Lemma A.4. *The function $x \mapsto \mathrm{sinc}(x)$ has bounded quadratic variation.*

Proof. It is shown that $v_2(\mathrm{sinc}) < \infty$. For every $n \in \mathbb{N}$, the sum $\sum_{k=1}^n [\mathrm{sinc}(x_k) - \mathrm{sinc}(x_{k-1})]^2$ is maximum for $x_k = (2k+1)\pi/2$, $k = 1, \ldots, n$. Splitting the sum into two parts,

$$\sum_{1 \leq k = 2j \leq n} [\mathrm{sinc}(x_k) - \mathrm{sinc}(x_{k-1})]^2 = \frac{4}{\pi^2} \sum_{1 \leq 2j \leq n} \left[\frac{8j}{(4j+1)(4j-1)}\right]^2$$

and

$$\sum_{1 \leq k = 2j+1 \leq n} [\mathrm{sinc}(x_k) - \mathrm{sinc}(x_{k-1})]^2 = \frac{4}{\pi^2} \sum_{1 \leq 2j+1 \leq n} \left[\frac{4(2j+1)}{(4j+3)(4j+1)}\right]^2.$$

Thus, $v_2(\mathrm{sinc}) < \infty$ as a consequence of $\sum_{j=1}^\infty j^{-2} < \infty$. $\square$

The following lemma provides an upper bound on the L^p-norm, $1 \leq p < 2$, in terms of the product of the L^∞-norm and any L^q-norm, $q > 1$. The proof is similar to that of statement (b) of Lemma 4 by Nguyen [25], pages 18 and 24.

LEMMA A.5. *Let $f,\, g \in L^\infty(\mathbb{R})$ be probability densities with $\mathrm{E}_f[|X|^u] < \infty$ and $\mathrm{E}_g[|X|^u] < \infty$ for some real $u > 0$. For every $1 \le p < 2$ and $t > 0$ such that $pt > 1$,*

$$\|f-g\|_p^p \le (s^{-1}+u) \\ \times \{s^{-1/s}(2^{1/s}/u)^u \|f-g\|_{pt}^{pu}\|f-g\|_\infty^{(p-1)/s}(\mathrm{E}_f[|X|^u]+\mathrm{E}_g[|X|^u])^{1/s}\}^{s/(1+su)},$$

where $s^{-1} := 1 - t^{-1}$.

PROOF. For every $R > 0$, by Hölder's inequality, $\int_{|x|\le R} |f(x)-g(x)|^p \,\mathrm{d}x \le (2R)^{1/s}\|f-g\|_{pt}^p$. Also, $\int_{|x|>R} |f(x)-g(x)|^p \,\mathrm{d}x \le R^{-u}\|f-g\|_\infty^{p-1}(\mathrm{E}_f[|X|^u] + \mathrm{E}_g[|X|^u])$. Thus, $\|f-g\|_p^p \le \inf_{R>0}\{(2R)^{1/s}\|f-g\|_{pt}^p + R^{-u}\|f-g\|_\infty^{p-1}(\mathrm{E}_f[|X|^u] + \mathrm{E}_g[|X|^u])\}$. The inequality in the assertion follows from $\min_{x>0}(Ax^\alpha + Bx^{-\beta}) = (\alpha+\beta)[(A/\beta)^\beta(B/\alpha)^\alpha]^{1/(\alpha+\beta)}$ for every $A,\, B,\, \alpha,\, \beta > 0$. □

The following lemma provides an upper bound on the number of components of a mixture, whose kernel density has Fourier transform satisfying (2.1), which uniformly approximates a given compactly supported mixture with the same kernel.

LEMMA A.6. *Let K be a probability density with characteristic function satisfying (2.1) for some constants $\rho,\, r > 0$ and $0 < L < \infty$. Let $0 < \varepsilon < 1$, $0 < a < \infty$ and $\sigma > 0$ be given. For any probability measure F on $[-a,\, a]$, there exists a discrete probability measure F' on $[-a,\, a]$, with at most*

$$N \lesssim \max\{\log(1/\varepsilon),\, (a/\sigma)\}, \qquad \text{if} \quad S_K < \infty,$$

and

$$N \lesssim \begin{cases} \log(1/\varepsilon), & \text{if} \quad 0 < r < 1 \\ & \text{and} \;\; \rho\sigma/a = O((\log(1/\varepsilon))^{(1-r)/r}), \\ \log(1/\varepsilon), & \text{if} \quad r = 1 \quad \text{and} \quad a/(\rho\sigma) \le e^{-1}, \\ \max\left\{\log(1/\varepsilon),\, (a/\sigma)^{r/(r-1)}\right\}, & \text{if} \quad r > 1 \quad \text{and} \quad a/(\rho\sigma) \ge e^{-1}, \end{cases}$$

if $S_K = \infty$, support points, such that $\|F * K_\sigma - F' * K_\sigma\|_\infty \lesssim \varepsilon/\sigma$.

PROOF. By Lemma A.1 of Ghosal and van der Vaart [10], page 1260, there exists a discrete probability measure F' on $[-a,\, a]$, with at most $N+1$ support points, N being a positive integer to be suitably chosen later on, such that it matches the (finite) moments of F up to the order N,

$$\mathrm{E}_{F'}[\Theta^j] := \int_{-a}^a \theta^j \,\mathrm{d}F'(\theta) = \int_{-a}^a \theta^j \,\mathrm{d}F(\theta) =: \mathrm{E}_F[\Theta^j],\;\; j = 1,\, \ldots,\, N. \tag{A.11}$$

By the moment matching condition (A.11),

$$|\hat{F}(t)-\widehat{F'}(t)| \leq \int_{-a}^{a} \frac{|t\theta|^N}{N!} \min\left\{\frac{|t\theta|}{N+1},\, 2\right\} \mathrm{d}(F+F')(\theta), \qquad t \in \mathbb{R}, \tag{A.12}$$

where the inequality holds because F and F' have finite absolute moments of any order, see, *e.g.*, inequality (26.5) in Billingsley [2], page 343. By (2.1), $\int_{\mathbb{R}} |\hat{K}(\sigma t)|\,\mathrm{d}t < \infty$, hence $F*K_\sigma$ and $F'*K_\sigma$ can be recovered using the inversion formula. By (A.12), $\|F*K_\sigma - F'*K_\sigma\|_\infty \leq 2a^N/(\pi N!)\int_{\mathbb{R}} |t|^N |\hat{K}(\sigma t)|\,\mathrm{d}t$. Next, we distinguish the case where $S_K < \infty$ from the case where $S_K = \infty$. If $S_K < \infty$, by the assumption that K satisfies (2.1),

$$\begin{aligned}\|F*K_\sigma - F'*K_\sigma\|_\infty &\leq \frac{2}{\pi}\frac{a^N}{N!}\int_{|t|\leq S_K/\sigma} |t|^N |\hat{K}(\sigma t)|\,\mathrm{d}t \\ &\leq \frac{4}{\sigma}[L + C(\rho,\, r)/\pi]\left(\frac{aeS_K}{\sigma N}\right)^N \lesssim \frac{\varepsilon}{\sigma}\end{aligned}$$

for $N \gtrsim \max\left\{\log(1/\varepsilon),\, (ae^2 S_K/\sigma)\right\}$. If $S_K = \infty$, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\|F*K_\sigma - F'*K_\sigma\|_\infty &\leq \frac{2}{\pi}\frac{a^N}{N!}\left(\frac{2\pi L}{\sigma}\right)^{1/2}\left(\int_{\mathbb{R}} |t|^{2N} e^{-2(\rho\sigma|t|)^r}\,\mathrm{d}t\right)^{1/2} \\ &\lesssim \frac{1}{\sigma}\left(\frac{a}{2^{1/r}\rho\sigma}\right)^N \frac{[\Gamma((2N+1)/r)]^{1/2}}{\Gamma(N+1)}.\end{aligned}$$

Using the formula $\Gamma(az+b) \sim (2\pi)^{1/2} e^{-az}(az)^{az+b-1/2}$ ($z \to \infty$ in $|\arg z| < \pi$, $a > 0$),

$$\|F*K_\sigma - F'*K_\sigma\|_\infty \lesssim \frac{1}{\sigma}\left(\frac{a}{\rho\sigma}\right)^N e^{N(1-1/r)} r^{-N/r} N^{-N(1-1/r)+(1/r-3/2)/2}.$$

If $0 < r < 1$ and $(\rho\sigma/a)^{r/(1-r)} = O(\log(1/\varepsilon))$, for

$$\left(\log\frac{1}{\varepsilon}\right) \lesssim N \lesssim \left(\frac{\sigma}{a}\right)^{r/(1-r)},$$

we have

$$\begin{aligned}&\|F*K_\sigma - F'*K_\sigma\|_\infty \\ &\quad\lesssim \frac{1}{\sigma}N^{(1/r-3/2)/2}\exp\left\{-N\left[\log\frac{\rho\sigma/a}{N^{1/r-1}} - \left(1-\frac{1}{r}+\frac{1}{r}\log\frac{1}{r}\right)\right]\right\} \lesssim \frac{\varepsilon}{\sigma}.\end{aligned}$$

If $r=1$ and $a/(\rho\sigma)\le e^{-1}$, for $N=\log(1/\varepsilon)$,

$$\|F*K_\sigma - F'*K_\sigma\|_\infty \lesssim \frac{1}{\sigma}\left(\frac{a}{\rho\sigma}\right)^N \lesssim \frac{\varepsilon}{\sigma}.$$

If $r>1$ and $a/(\rho\sigma)\ge e^{-1}$, for

$$N \lesssim \max\left\{\left(\log\frac{1}{\varepsilon}\right),\, \left(\frac{a}{\sigma}\right)^{r/(r-1)}\right\},$$

we have

$$\begin{aligned}\|F*K_\sigma - F'*K_\sigma\|_\infty &\\ &\lesssim \frac{1}{\sigma}\exp\left\{-N\left[\log\frac{N^{1-1/r}}{a/(\rho\sigma)}-\frac{1}{r}(r-1-\log r)\right]\right\}\lesssim\frac{\varepsilon}{\sigma}\end{aligned}$$

and the proof is complete. □

REMARK A.2. Even if stated for a probability measure F supported on a symmetric interval $[-a,\, a]$, Lemma A.6 holds for every F with $\mathrm{supp}(F)$ being *any* compact interval.

LEMMA A.7. *Let f and g be probability densities on $\mathbb{R}$. For every $\upsilon\in(0,\,1]$ such that $\int f^\upsilon\,\mathrm{d}\lambda<\infty$, we have $\|f-g\|_1\le 2\|f-g\|_\infty^{1-\upsilon}\int f^\upsilon\,\mathrm{d}\lambda$.*

PROOF. Write $\|f-g\|_1 = 2\int(f-g)_+\,\mathrm{d}\lambda \le 2\int\min\{f,\,\|f-g\|_\infty\}\,\mathrm{d}\lambda \le 2\|f-g\|_\infty^{1-\upsilon}\int f^\upsilon\,\mathrm{d}\lambda$. The assertion follows. □

As noted in Devroye [5], Remark 3, page 2042, if

$$\text{for some } q>0,\qquad \mathrm{E}_f[|X|^q]<\infty, \tag{A.13}$$

then $\int f^\upsilon\,\mathrm{d}\lambda<\infty$ whenever $\upsilon>(1+q)^{-1}$. For example, condition (A.13) is verified for a Student's-t distribution with ν degrees of freedom when $0<q<\nu$.

The inequality of the following lemma can be proved similarly to the one for the Gaussian kernel, see, *e.g.*, the first part of Lemma 1 in Ghosal *et al.* [8], pages 156–157.

LEMMA A.8. *Let K be a probability density on $\mathbb{R}$, bounded and symmetric around 0. For every $\sigma>0$ and every $\theta_j,\,\theta_k\in\mathbb{R}$,*

$$\|K_\sigma(\cdot-\theta_j)-K_\sigma(\cdot-\theta_k)\|_1 \,\le\, 2\,\|K\|_\infty\frac{|\theta_j-\theta_k|}{\sigma} \,\lesssim\, \frac{|\theta_j-\theta_k|}{\sigma}.$$

In the next lemma, a sufficient condition is provided for the L^1-distance between kernel mixtures with different variances to be bounded above by the distance between the variances.

LEMMA A.9. *Let K be a probability density on $\mathbb{R}$ symmetric around 0 and monotone decreasing in $|x|$. For every probability measure F on $\mathbb{R}$ and every $\sigma, \sigma' > 0$, we have $\|F * K_\sigma - F * K_{\sigma'}\|_1 \le \|K_\sigma - K_{\sigma'}\|_1 \le 2|\sigma - \sigma'|/(\sigma \wedge \sigma')$.*

PROOF. Note that

$$\|F * K_\sigma - F * K_{\sigma'}\|_1 \le \int_{\mathbb{R}} \|K_\sigma(\cdot - \theta) - K_{\sigma'}(\cdot - \theta)\|_1 \, \mathrm{d}F(\theta) = \|K_\sigma - K_{\sigma'}\|_1.$$

The second inequality in the assertion can be proved as in Norets and Pelenis [26], page 18. □

The next lemma provides an upper bound on the L_1-metric entropy of sets of mixtures with super-smooth kernels. It is based on Lemma A.6, Lemma A.8 and Lemma A.9 and can be proved similarly to Lemma 3 of Ghosal and van der Vaart [12], pages 705–707, which deals with normal mixtures.

LEMMA A.10. *Let K be a probability density on $\mathbb{R}$ symmetric around 0 and monotone decreasing in $|x|$, with characteristic function satisfying (2.1) for some constants $\rho, r > 0$ and $0 < L < \infty$. Let $0 < \varepsilon < 1/5$. Let $0 < s \le S < \infty$ and $0 < a < \infty$ be such that, for some $\nu > 0$, $(a/s) \lesssim (\log(1/\varepsilon))^\nu$. Define $\mathscr{F}_{a,s,S} := \{F * K_\sigma : F([-a, a]) = 1, \;\; s \le \sigma \le S\}$. Then,*

$$\log N(\varepsilon, \mathscr{F}_{a,s,S}, \|\cdot\|_1) \lesssim \log\left(\frac{S}{s\varepsilon}\right) + N \times \left[\log\left(\frac{2a}{s\varepsilon} + 1\right) + \log\frac{1}{\varepsilon}\right],$$

where

$$N \lesssim \begin{cases} \dfrac{a}{s} \times \left(\log\dfrac{1}{\varepsilon}\right)^{1/r}, & \text{if } \; 0 < r \le 1, \\[2ex] \left[\left(\dfrac{a}{s}\right)^{r/(r-1)} \vee \left(\log\dfrac{1}{\varepsilon}\right)\right], & \text{if } \quad r > 1. \end{cases}$$

The following lemma is a slight variant of Lemma 6 in Ghosal and van der Vaart [12], page 711.

Lemma A.11. *Let K be a probability density on $\mathbb{R}$ symmetric around 0. Let f be a strictly positive and bounded probability density, non-decreasing on $(-\infty,\,a)$, non-increasing on $(b,\,\infty)$ and such that $f(x)\geq \ell>0$ on $[a,\,b]$. For every $\zeta\in(0,\,1)$, let $\tau_\zeta>0$ be such that $\int_0^{b-a}K_{\tau_\zeta}(x)\,\mathrm{d}x\geq\zeta$. Then, for every $\sigma<\tau_\zeta$, we have $f*K_\sigma\geq C_\zeta f$, with $C_\zeta:=(\zeta\ell/\|f\|_\infty)\in(0,\,1)$.*

ACKNOWLEDGEMENTS

The author is grateful to Prof. A. van der Vaart for his availability, insightful comments, remarks and suggestions while visiting the Department of Mathematics, Faculty of Sciences, VU University, Amsterdam, whose kind hospitality is gratefully acknowledged.

REFERENCES

[1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Applied Mathematics Series, **55**. U.S. Government Printing Office, Washington, D.C. Available online at http://www.math.sfu.ca/ cbm/aands/

[2] Billingsley, P. (1995). *Probability and Measure*, 3rd ed., John Wiley & Sons, Inc., New York.

[3] Butucea, C. and Tsybakov, A. B. (2008). Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52** 24–39.

[4] de Jonge, R. and van Zanten, J. H. (2010). Adaptive nonparametric Bayesian inference using location-scale mixture priors. *Ann. Statist.* **38** 3300–3320.

[5] Devroye, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20** 2037–2056.

[6] Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, Eds.) 287–302. Academic Press, New York.

[7] Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29** 1264–1280.

[8] Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158.

[9] Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531.

[10] Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263.

[11] Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35** 192–223.

[12] Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.* **35** 697–723.

[13] Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in L^r-metrics, $1\leq r\leq\infty$. *Ann. Statist.* **39** 2883–2911.

[14] Hasminskii, R. and Ibragimov, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18** 999–1010.

[15] Hurst, S. (1995). The characteristic function of the Student t distribution. *Financial Mathematics Research Report* No. FMRR006-95, *Statistics Research Report* No. SRR044-95.

[16] Ibragimov, I. (2001). Estimation of analytic functions. In *State of the Art in Probability and Statistics. Festschrift for W. R. van Zwet*, 359–383. IMS, Hayward, CA.

[17] Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* **87** 371–390.

[18] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173.

[19] Kawata, T. (1972). *Fourier Analysis in Probability Theory*. Probability and Mathematical Statistics, No. **15**. Academic Press, New York-London.

[20] Kruijer, W., Rousseau, J. and van der Vaart, A. (2010). Adaptive Bayesian density estimation with location-scale mixtures. *Electron. J. Statist.* **4** 1225–1257.

[21] Lijoi, A., Mena, R. H. and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100** 1278–1291.

[22] Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12** 351–357.

[23] Maugis, C. and Michel, B. (2011). Adaptive density estimation using finite Gaussian mixtures. `arXiv:1103.4253`

[24] Mena, R. H., Ruggiero, M. and Walker, S. G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling. *J. Statist. Plann. Inference* **141** 3217–3230.

[25] Nguyen, X. (2011). Convergence of latent mixing measures in nonparametric and mixture models. *Preprint*.

[26] Norets, A. and Pelenis, J. (2011). Posterior consistency in conditional density estimation by covariate dependent mixtures. *Preprint*.

[27] Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900.

[28] Pogány, T. K. (2010). Closed expression for characteristic function of CEPE distribution. *J. Math. Res.* **2** 49–50.

[29] Pogány, T. K. and Nadarajah, S. (2010). On the characteristic function of the generalized normal distribution. *C. R. Acad. Sci. Paris, Ser. I* **348** 203–206.

[30] Scricciolo, C. (2011). Posterior rates of convergence for Dirichlet mixtures of exponential power densities. *Electron. J. Statist.* **5** 270–308.

[31] Shen, W. and Ghosal, S. (2011). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. `arXiv:1109.6406`

[32] Titchmarsh, E. C. (1937). *Introduction to the theory of Fourier integrals*. The Clarendon press (Oxford).

[33] Wong, W. H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362.

Department of Decision Sciences
Bocconi University
Via Röntgen 1, I-20136 Milan, Italy
E-mail: catia.scricciolo@unibocconi.it